”Who’s the criminal?“

Early detection of hidden criminal intentions - Influence of nonverbal behavioral cues, theoretical knowledge, and professional experience

Thesis

Presented to the Faculty of Arts and Social Sciences

of the University of Zurich

for the Degree of Doctor of Philosophy

by

Corinne Ines Frey

Accepted in the Autumn Term 2014

on the Recommendation of the Doctoral Committee:

Prof. Dr. Klaus Oberauer (main advisor)

Prof. Dr. Claus-Christian Carbon

Zurich, Switzerland, 2014

Abstract

Prevention of criminal activities, especially but not only of possible terrorist attacks, is of high importance for society. The attack on the World Trade Center in the U.S. on September 11, 2001, evolved in high investments towards aviation security. Several new technologies, such as liquid explosives detection systems and body scanners were implemented. However, apart from new technologies, human factors were also taken into account: The development of training and selection tools for baggage screeners as well as humans' abilities to infer other's intentions based on nonverbal behavioral cues were considered as possible prevention tool. So-called *behavior detection programs* were implemented, which are based on the assumptions that behavior detection officers are able to infer hidden criminal and terroristic intentions during the build-up phase of the criminal or terroristic act, using nonverbal behavior as cues. The aim of this thesis was to investigate a) whether such hidden criminal intentions during the build-up phase are recognizable in nonverbal behavioral cues, b) whether observers are able to detect such intentions during the build-up phase based on nonverbal behavioral cues, and c) whether the detection of hidden criminal intentions is trainable.

In the first reported study, five groups of participants (students of psychology, police recruits, inexperienced police officers, experienced police officers and criminal investigators) were examined in order to investigate whether hidden criminal intentions can be detected during the build-up phase, and whether theoretical knowledge and work experience influence this detection performance positively. To this end, all participants answered a computer-based test containing authentic CCTV recordings of the build-up phase of real baggage thefts at a large international airport. The results indicate that shortly before the theft happens, detection of hidden criminal intention based on nonverbal behavioral cues is possible even for laypeople. Nevertheless, theoretical knowledge alone already improves detection performance

impressively at all three measured points during the build-up phase. Special work experience of criminal investigators then improves performance even more. Thus, the results of this study indicate that the detection of hidden criminal intentions during the build-up phase based on nonverbal behavioral cues is indeed possible and that it might be trainable.

The second study was then conducted to address the question which nonverbal behavioral cues are relevant for the detection of hidden criminal intentions during the build-up phase, thus allowing such performance as observed in the first study. In order to control for inter-individual differences, own recordings of the same people in a search (control) condition and a mock crime (test) condition were created. Two groups of participants who were naïve about the actual goal of the study analyzed these recordings according to the behaviors of interest (Experiments 1 and 2). Experiment 1 was conducted to investigate differences in *moving patterns* in public spaces and *communication behavior* between offenders and non-offenders. Additionally to the created recordings, ten authentic CCTV recordings from the first study were used to test ecological validity. The observed behaviors of interest were compared to the baseline (i.e. how strongly this behavior is expected to be expressed from people without hidden criminal intentions in the same situation) as well as between the three conditions (search, mock crime, and real crime). In Experiment 2, *self-* and *object-adaptors* between the search and the mock crime conditions were examined as the behaviors of interest. The results of both experiments revealed significant differences in offender and non-offender behavior. Moreover, they clearly pointed to the relevance of the construct baseline and the need to judge clusters of nonverbal behavior such as *moving patterns* instead of focusing on individual explicit behaviors such as *abrupt changes in direction*.

Taken together, the results of both studies reported in this thesis clearly indicate that the implementation of behavior detection programs might be a reasonable way of crime-prevention programs.

Zusammenfassung

Die Prävention von kriminellen und insbesondere terroristischen Aktivitäten ist für die Sicherheit der Gesellschaft von grosser Bedeutung. Nach dem Anschlag auf das World Trade Center in den USA am 11. September 2001 wurde viel in die Verbesserung der Flughafensicherheit investiert. Neben der Implementierung neuer Technologien wie beispielsweise Detektionsgeräte für Flüssigsprengstoffe oder Ganzkörperscanner wurde dabei auch vermehrt Aufmerksamkeit auf die sogenannten *human factors* gerichtet. Dies führte zur Entwicklung und Einführung von Auswahl- sowie Trainingsverfahren für Gepäckkontrollpersonal. Zudem wurden sogenannte Verhaltenserkennungsprogramme eingeführt. Diese basieren auf der Annahme, dass kriminelle oder terroristische Intentionen bereits vor der eigentlichen Tat anhand des nonverbalen Verhaltens erkannt werden können. Das Ziel dieser Dissertation war, wissenschaftlich zu überprüfen, ob versteckte kriminelle Intentionen während der Vortatphase tatsächlich im nonverbalen Verhalten sichtbar werden und ob Beobachter auch wirklich in der Lage sind, diese zu erkennen. Zusätzlich sollten erste Hinweise auf eine mögliche Lern- und Trainierbarkeit dieser Verhaltenserkennung gefunden werden.

Um herauszufinden, ob versteckte kriminelle Intentionen während der Vortatphase erkannt werden können und welchen Einfluss Theorie- und Erfahrungswissen auf diese Erkennungsleistung haben, wurden in der ersten Studie fünf Gruppen von Teilnehmenden untersucht (Psychologiestudierende, Polizeirekruten, unerfahrene Polizisten, erfahrene Polizisten und Kriminalpolizisten). Alle Gruppen absolvierten einen eigens dafür entwickelten computerbasierten Test, der authentisches Videomaterial von Überwachungskameras enthielt, welches die Vortatphase erfolgreicher Gepäckdiebstähle zeigte. Die Resultate der Studie zeigen, dass das Erkennen von kriminellen Intentionen während der Vortatphase am nonverbalen Verhalten tatsächlich möglich ist. Wie der

Messzeitpunkt kurz vor der Tat zeigte, sind bereits Laien (Psychologiestudierende und Polizeirekruten) dazu fähig kriminelle Intentionen zu erkennen. Das theoretische Wissen von unerfahrenen Polizisten führte zu einer deutlichen Verbesserung der Erkennungsleistung zu allen drei Messzeitpunkten während der Vortatphase, welche nur durch das explizite Erfahrungswissen von Kriminalpolizisten übertroffen wurde. Diese Ergebnisse stärken die Annahme der Trainier- und Lernbarkeit dieser Fähigkeit.

Die zweite Studie widmete sich der Frage, welche nonverbalen Verhaltensweisen während der Vortatphase konkret Hinweise auf versteckte kriminelle Intentionen liefern. Um inter-individuelle Unterschiede im nonverbalen Verhalten zu kontrollieren, wurden eigene Videoaufnahmen von denselben Personen in einer Kontrollsuchbedingung und einer gestellten Tatbedingung erstellt. Zur Überprüfung der ökologischen Validität dieses Videomaterials wurden zehn Videos aus der ersten Studie mitanalysiert. Zwei Laiengruppen, die nicht über den Hintergrund der Studie informiert waren, analysierten die Videos anhand der vorgegebenen Verhaltensweisengruppen (Experiment 1 und Experiment 2). In Experiment 1 wurden die Verhaltensweisengruppen *Bewegungen im Raum* und *Kommunikations-verhalten* genauer untersucht. In Experiment 2 wurde auf die Verhaltensweisengruppen *Objekt-* und *Selbstadaptoren* fokussiert. Die Ergebnisse dieser Studie zeigen, dass es beim Vergleich des nonverbalen Verhaltens zwischen Tätern und Nichttätern Unterschiede gibt. Zusätzlich zeigte sich anhand der Daten die Wichtigkeit des Vergleichs mit einer Baseline (Wie stark kann erwartet werden, dass dieses Verhalten von anderen Personen ohne Tatabsicht in derselben Situation am selben Ort gezeigt wird?). Zudem ist es vorteilhafter, Cluster von Verhaltensweisen wie beispielsweise *Bewegungen im Raum* anstelle von expliziten einzelnen Verhaltensweisen (z.B. *abrupte Richtungswechsel*) zu beurteilen.

Die Ergebnisse der Studien, die in dieser Dissertation dargestellt werden, deuten darauf hin, dass sich Verhaltenserkennungsprogramme zur Unterstützung von Präventionsarbeit gegen kriminelle Aktivitäten eignen.

Content

1 Introduction

1.1 Background

After the attack on September, 11, 2001, (9/11) when four hijacked airplanes crashed into several important buildings on U.S. grounds, the public was paralyzed (BBC News, 2001). This horrendous, but at the same time impressively planned, attack to destroy thousands of lives as well as buildings representing national security of the U.S. (pentagon) and economic influence and stability (World Trade Center) resulted in an evolvement of the investments into counter terrorism and therefore aviation security (Clarke & Newman, 2006) – thus, representing the typical action-response-reaction relationship within the security branch (Mitchener-Nissen, Bowers, & Chetty, 2012). Especially in aviation security all new amendments were a reaction to intelligence-service knowledge about a planned or actual attack (Mitchener-Nissen et al., 2012), and every new amendment forced attackers to change their modus operandi (Clarke & Newman, 2006).

In general, two possible ways to improve security can be distinguished: a) technological improvements and b) improvements concerning the so-called “human factors”. Mitchener-Nissen et al. (2012) describe impressively how the improvement of aviation security technologies happened over the last forty decades. Since the 1970s, planes were successfully attacked by means of smuggled weapons and explosives in hand baggage. As a response, aviation security introduced metal detectors as well as X-ray screening of hand baggage, which then forced the attackers to switch weapons and explosives into checked baggage. Thus, law enforcement started to not let passengers with suspicious baggage enter the plane. This resulted in a decreasing rate of attacks. The reaction of attackers to this new kind of aviation security resulted in the new modus operandi now very well known by the public as the horrendous attacks on 9/11, where hijackers turned the plane itself into a

gigantic explosive (National Commission on Terrorist Attacks Upon the United States, 2004). Since then, suicide bombings have become one of the greatest threats to aviation security (Hofer & Wetter, 2012). Due to the incapability of metal detectors to signal organic material and the former allowance of fluids in hand baggage, suicide bombers could transport liquid and/or organic explosives on body or in hand baggage, e.g. Richard Reid also known as the shoe bomber (Frank, Maccario, & Govindaraju, 2009; Mitchener-Nissen et al., 2012).

After 9/11 and the prevention of planned attacks as the above mentioned shoe bombing, several new kinds of technologies were developed, for example, liquid explosives detection systems (LEDS: Hofer & Wetter, 2012) and body scanners (Mock, 2009; Bart, 2012; Hofer & Wetter, 2012; Mitchener-Nissen et al., 2012). The combination of these two technologies should make it impossible for attackers to enter an airplane with organic explosives.

However, the best technology is just as good as the human that operates it. For example, it is known that the 9/11 hijackers could enter the planes with knives and box cutters, indicating the inadequacy of hand baggage screening at that time (Mitchener-Nissen et al., 2012). This example dramatically shows the high importance of human factors in security, because the visual inspection of hand baggage to find threat objects using X-ray imaging is one of the most important tasks of security personnel at airports (Hardmeier, Hofer, & Schwaninger, 2005). It is important that the operating humans are selected and trained to perform their task efficiently, otherwise the technology is not of much value (Schwaninger, 2005).

In general, object-recognition happens when the perceived visual information matches the visual representation stored in memory. The stored representation comes from prior exposure to the respective stimuli (Schwaninger, 2005). According to Schwaninger (2005) the most important factors influencing the detection performance of a baggage screener are a) the

knowledge about prohibited items and b) the knowledge about what they look like in X-ray images. Security personnel might have rather good knowledge about which items are prohibited and they might also have a solid representation of how these items look in real life. But as baggage screeners at an airport, they have to be able to recognize these items in an X-ray image and in a fully packed baggage. As a consequence, they must be able to detect such items at different positions in different orientations and (usually) overlapped by other objects (Schwaninger, Hardmeier, & Hofer, 2004). Back to the example of the 9/11 hijackers, if screeners have never seen what a box cutter looks like in an X-ray image, they will not be able to detect it as a prohibited item. Or if they have never seen a knife from the backside, chances are high that they will not be able to identify it as a threat. It was therefore important for an increase of aviation security to develop a test and training tool which allowed to select and train security baggage screener in detecting threat items in hand baggage (Schwaninger & Hofer, 2004; Hardmeier et al., 2005; Schwaninger, 2005). According to Schwaninger (2005), the "X-Ray Tutor", a training tool for baggage screener, encountered many different airports all over Europe and U.S. since 2002.

Despite the fact that such selection and training tools show promising improvements, they too have to be updated continuously as reaction to new threats. When British law enforcement arrested several people on August 9, 2006, they found strong indications for the plan to blow up several aircrafts by means of liquid explosives (Schwaninger, 2006). Consequently, liquids were prohibited in hand baggage and baggage screeners needed to be trained in detecting liquids in X-ray images.

LEDS, body scanners, X-ray imaging for hand baggage, improving all these technologies is important but the use of all these parts reduces passenger flow at an airport remarkably (Hofer & Wetter, 2012). Thus, how much more convenient would it be to catch

possible criminals and terrorists during the build-up phase of their act with only human observers with as less influence on passenger flow as possible?

1.1.1 Behavior detection and its role in security

The idea of early-on identification of criminal/terroristic intentions was implemented in the development of so-called behavior detection programs (BD programs) which are based on the reasoning that it should be possible to infer criminal intentions based on nonverbal behavior (U.S. Government Accountability Office, 2010). Israel (*Behavior Pattern Detection*) and the U.S. (*SPOT*: *S*creening *P*assengers by *O*bservation *T*echnique) were the pioneering countries concerning such behavior detection programs (Frank et al., 2009; U.S. Government Accountability Office, 2010). Meanwhile several European countries implemented similar programs, e.g. U.K.'s Griffin Project (City of London Police, 2014). As an example, *SPOT* trained officers observe the passengers based on a checklist consisting of mostly nonverbal behavioral cues. According to U.S. Government Accountability Office (2010) each of these nonverbal behavioral cues has an according amount of points. As soon as the behavior of passengers exceeds baseline behavior (behavior that is typically expected in such a situation) observers must mentally add up the according points of the observed behaviors. If the summed up points reach a certain threshold, the officer hands the passenger over to the police for further investigation (U.S. Government Accountability Office, 2010). The importance of such programs – if effective – is quite obvious; they could also be implemented in daily work of law enforcement not only in counter terrorism.

The implementation of behavior detection programs all over the world indicates the high expectations and beliefs towards inferring other's intentions from nonverbal behavior. Already 40 years ago, many journalists started to write popular books on nonverbal communication which mostly over-interpreted what actually could be concluded from research (Koivumaki, 1975; Masip, 2005). Probably the most famous research finding often

mistakenly described in such popular books is the so-called *mehrabian formula*, based on the studies of Mehrabian and Wiener (1966) and Mehrabian and Ferris (1967). The formula states that only 7% of a message is sent via the actual spoken content and 93% via nonverbal channels (e.g. mimics and pitch). Mehrabian and colleagues correctly describe in their studies that this formula was based on the communication of positive and negative emotions by using only ONE word. In some popular books authors erroneously generalize this finding on whole messages and argue that nonverbal behavior as a whole is the most important channel for all communication (e.g. Kuhnke, 2008). In order to interpret nonverbal behavior, authors propose rather bogus statements (Hecht & Ambady, 1999), such as "crossing the legs or arms clearly signifies that the person is not psychologically "open" to the other" (Masip, 2005, p. 79). These untested beliefs of the authors of such books had enormous influence on interested but unfamiliar readers, consequently leading them to wrongly attributing personality traits to people, for example in a job interview. In extreme cases this might even result in wrongly reject a person (Masip, 2005).

Research on nonverbal communication mostly contradicts the public's beliefs, nevertheless, it seems as if these stereotypes had more value outside of the scientific community due to the more popularity of such "self-help" books (Masip, 2005). As Masip (2005) points out, this is even more relevant considering the detection of lies based on nonverbal behavior. If a suspect was interrogated about a murder and the interrogator relied on nonverbal behavioral cues to deception proposed in such books, the life of the suspect could be at stake in case the interrogator's statement about the credibility of the suspect is taken as evidence in a court trial (Masip, 2005; Vrij, 2008). Inbau, Reid, Buckley, and Jayne (2001), for example, published a training program designed for security personnel who had to interrogate suspects and witnesses. As a part of this program deception cues are taught (Vrij, 2008). Kassin and Fong (1999) showed empirically that people, trained by means of the Inbau et al. (2001) cues to deception, performed worse after the training and were biased towards

identifying lies. Moreover, they were even more confident of their statements. These results do not surprise if research about cues to deception is considered, as the cues trained could not be confirmed by empirical research on deception (Masip, 2005). As a consequence, this training does not only influence the performance in lie detection, it also decreases performance in judging truth statements (Mann, Vrij, & Bull, 2004). According to Masip (2005), more than 300,000 professionals have been trained with Inbau's et al. (2001) training. This example dramatically demonstrates the danger of such programs if they are not scientifically examined carefully. Despite that the Inbau training focuses on interrogation settings, similar false beliefs presumably also exist for the expressed nonverbal behavior due to an underlying hidden criminal intention. Therefore, for a behavior detection program as SPOT to be effective, three conditions must be met: First, offenders' *hidden* criminal intentions must be displayed in nonverbal behavior that triggers suspiciousness. Second, observers must be able to infer *hidden* criminal intentions based on nonverbal behavioral cues, and, third, the detection of *hidden* criminal intentions during the build-up phase must be teachable and trainable.

Before any of these conditions can be accepted, one general underlying presumption must be met, namely that our intentions are expressed in nonverbal behavior *at all*. Considering our daily life it is reasonable to assume that we express intentions via the nonverbal channel, and that we are able to infer other's intentions based on nonverbal behavior. For example, if we saw somebody running towards a waiting train waving his/her hands we would almost immediately interpret the running behavior as the expressed intention of reaching this train and the waving as asking us to keep the door from closing. Thus, some of our intentions are indeed expressed in nonverbal behavior and at least if the behavior is shown deliberately (e.g. running and waving) we are able to infer the underlying intention correctly (see Baldwin & Baird, 2001, for review).

1.1.2 Hidden intentions expressed in nonverbal behavior

Already Darwin (1872/1998) postulated that besides emotions mimics might also express thoughts or intentions. Regarding the expression of emotions in mimics he even proposed that these were universal. If all people would express emotions differently there would be no possibility to define cues to hidden criminal intentions during the build-up phase, because every offender would express things completely different. Emotional expressions being universal, seems to be rather trivial today, but at the time Darwin sketched his ideas it was not as he used it to support his *evolution theory*.

Ekman and Friesen (1969b) and Ekman et al. (1987) showed scientifically that at least facial expressions of basic emotions (e.g. fear, happiness) are indeed universal. However, Ekman (1998) also showed that Japanese students *masked* their felt emotions by smiling more, because in their culture showing negative emotions was not approved. Thus, culture and society codetermine the rules that manage facial expressions. This so-called *display rules* indicate that people might be able to control their behavior to some extent (Ekman, 1998). More important, people really try to control their behavior under certain circumstances (J. K. Burgoon & Buller, 1994), and they probably try so especially if they want to hide what they intend to do. Imagine a terrorist entering an airport having the intention to blow up a plane by suicide bombing. This terrorist will try to deceive security personnel, other passengers, or the flight crew. In other words, this terrorist will try to hide the intention and behave as inconspicuously as possible.

Ekman (2009) proposed that guilt, fear and delight were the emotions that underlie lies or deception. Although, these results come from lie detection research and are based mostly on interrogation settings, one could assume that similar emotions underlie a *hidden* criminal intention during the build-up phase. Therefore, concerning the imagined terrorist from above, on the one hand he/she could feel guilty for intending to destroy as many lives to achieve

his/her higher goal. On the other hand, he/she could be scared of being detected and not achieve the higher goal. At some point during the build-up phase, maybe shortly before entering the plane, he/she might also be delighted by almost having achieved the goal. Thus, the combination of a stressful situation, trying to appear truthful, and control of the own behavior might result in informative cues making observers able to detect the deception.

However, several studies examining lie detection abilities showed rather modest detection rates independent of the "expertise" of the observers (DePaulo & Pfeifer, 1986; Vrij & Graham, 1997; Meissner & Kassin, 2002; Hartwig, Granhag, Strömwall, & Vrij, 2004; Bond & DePaulo, 2006; see Vrij, 2008 for review). Thus, neither do we seem to be very good at detecting lies based on nonverbal behavior at all, nor does professional experience with interrogations (e.g. as a police officer) help to improve detection rate. These results are rather discouraging concerning the second and third condition for the effectiveness of a behavior detection program as they refer to the detection ability and trainability of detection.

It is important to notice, though, that the existing literature about cues to deception and lie detection ability focuses on interrogation settings (see Sporer & Schwandt, 2007; Vrij, 2008 for reviews), which is different to the build-up phase of a criminal act in many ways. Firstly, the interrogation setting that is researched in most studies does not have a build-up phase, because as soon as the interrogation starts the intention to hide the truth is already expressed. In other words, the actual build-up phase would be the waiting time until the interrogation starts. Lie detection, therefore, is more similar to the detection of the committed criminal act. In contrast, thieves or terrorists intend to steal or blow-up something and must be detected before they express the intentions. Second, the term *interrogation* already implies that the observed suspect directly and overtly interacts with the interviewer and therefore tries to convince somebody. This is in contrast to the criminals during the build-up phase, because they do not interact with security personnel and do not try to convince somebody in the same

way as a lying suspect during interrogation. Third, also truth-tellers try to convince the interrogator and therefore might alter their nonverbal behavior (Strömwall, Hartwig, & Granhag, 2006; Sporer & Schwandt, 2007). If truth-tellers as well as liars alter their nonverbal behavior during an interrogation and the interviewer does not know who is telling the truth, he/she never has a solid and real guideline for nonverbal behavior expressed when telling the truth. In other words, interviewers in interrogation settings never have a reliable guideline to compare suspects' behavior to. In contrast, offenders hiding their criminal intention during a build-up phase of a criminal act might try to become one with the crowd and therefore alter their behavior. However, bystanders in this situation mostly just go their regular way without having the need to persuade somebody of something. Consequently, bystanders might be a perfect guideline for observers to compare the behavior of possible suspect to. Thus, it might be easier to detect criminal intentions during the build-up phase than detecting a lie during an interrogation. Some evidence for this assumption comes from a study of Troscianko et al. (2004): They showed that at least the prediction of aggressive, unlawful acts based on nonverbal behavior is possible, although they found no difference in performance between laypeople and CCTV operators and the results could not be replicated (Grant & Williams, 2011). Still, some preliminary studies with rather small sample sizes and/or methodological issues showed promising results regarding the ability to detect the *hidden* criminal intention based on nonverbal behavior during the build-up phase (Heubrock, Kindermann, Palkies, & Röhrs, 2009a; Baettig, Frey, & Hofer, 2011; Heubrock, 2011). Despite these few preliminary studies, though, there is still need to examine the *hidden criminal intention* detection during the *build-up phase* in better controlled studies with larger sample sizes.

Another problem that has been prevalent in previous studies – with the exception of the study of Eachus, Stedmon, and Baillie (2013) – concerns the fact that always the behavior of different people as offenders and non-offenders was compared. This is problematic because the nonverbal behavior of a generally calm person in the control situation could be compared

to the nonverbal behavior of a generally nervous person in the crime situation or vice versa. This would result in an overestimation of possible cues to hidden criminal intentions or an underestimation, respectively.

Based on the lack of sound scientific evidence on the topic of detecting hidden criminal intentions during the build-up phase of a criminal act, developers of behavior detection programs might rely either on the ambiguous literature from self-help books about lie detection or/and serious literature on lie detection. Thus, when examining the build-up phase, the results from lie detection research have to be taken into account. Additionally, the underlying emotions and the attempted behavioral control might still be the same for liars in interrogations and criminals during the build-up phase. In the end, this might result in similar behaviors but before accepting this conclusion, additional evidence is required.

Therefore, the main goal of this thesis was to contribute scientific evidence towards the effectiveness of behavior detection programs, such as SPOT, by examining the three mentioned conditions underlying the successful detection of *deceptive* offenders during the *build-up* of their criminal act.

1.2 Empirical studies

In the first study reported in this thesis (*"Who's the thief?" The influence of knowledge and experience on early detection of criminal intentions*), we examined whether offenders (in our case thieves) who hide their intention can be identified out of a rather large crowd of bystanders during the build-up phase of the theft. In other words, examine the second condition for the effectiveness of BD programs: observers must be able to infer *hidden* criminal intentions based on nonverbal behavioral cues. Additionally, we were interested in the influence of theoretical knowledge received during the first education as a police officer and of work experience as a police officer. Thus, examine the third condition for the

effectiveness of BD programs: the detection of *hidden* criminal intentions during build-up must be teachable and trainable. For this purpose, we examined five groups of participants (students of psychology, police recruits, inexperienced police officers, experienced police officers, and criminal investigators) by means of a specially developed computer-based test containing twelve authentic CCTV recordings of real baggage thefts at a large international airport. Students of psychology and police recruits represented laypeople, while inexperienced police officers were included to reveal the influence of theoretical knowledge. Experienced police officers and criminal investigators represented the influence of work experience per se and specific work experience in observing people, respectively. The participants were asked to identify thieves at three points during the build-up phase by a mouse click on the suspect(s) head(s). Data were analyzed according to Signal Detection Theory (Green & Swets, 1974). To foreshadow the main results, all groups of participants were able to identify thieves correctly shortly before the criminal act was actually committed. However, inexperienced police officers outperformed laypeople impressively already at the beginning of the build-up phase, thus pointing to the relevance of theoretical knowledge about offenders' *modi operandi* and the role of the context. Criminal investigators who were more specialized in observing people outperformed inexperienced police officers at every decision point during the build-up phase. Experienced police officers with no special experience in observing people, but on average having the same amount of work experience as a police officer as the criminal investigators, performed somewhere in between inexperienced police officers and criminal investigators. These latter results clearly point to the (additional) relevance of practical experience, and this practical experience helps the more the closer it is related to the actual task to be performed.

Based on the evidence from the first study that deceptive offenders, such as thieves, can indeed be detected during the build-up phase, we examined in the second study (*What is suspicious when trying to be inconspicuous? Criminal intentions inferred from nonverbal behavioral cues*) what possible nonverbal behavioral cues to hidden criminal intentions exist.

In other words, examine the first condition for the effectiveness of BD programs: offenders' *hidden* criminal intentions must be displayed in nonverbal behavior that triggers suspiciousness. In this study, we focused on a specific subset of visually recognizable nonverbal behavior: spatial moving patterns (e.g. how the person moves in a public space: Heubrock, 2011), communication behavior (e.g. hand signs: based on personal communication with police officers), self-adaptors (e.g. scratching or scrubbing the face with one hand: Sporer & Schwandt, 2007), and object-adaptors (e.g. use of an object like a cellphone without instrumental goal: Heubrock, Kindermann, Palkies, & Röhrs, 2009b).

For the purpose of this study, we created our own recordings additionally to the authentic CCTV recordings from the first study. These self-made recordings were taken from eight different mock offenders or teams of mock offenders. Each of the offenders first had to search a non-existing person. This task resulted in a goal-driven noncriminal intention and additionally in the possibility to later exclude mere search behavior from the build-up phase of a criminal act. After this task, they had to either try to steal something or place a mock bomb to "murder" as many people as possible. These tasks were recorded by a static camera filming the whole situation without ever zoom or move (*space* condition) and several cameras zooming into the actors (*zoom* condition).

Based on these recordings we conducted two experiments. In the first experiment we addressed *moving patterns in space* (e.g. abrupt changes in direction or speed) and *communication behavior* (e.g. hand signs, head movements) as the behaviors of interest. To validate our recordings ecologically, we used some of the authentic CCTV recordings from the first study. For the first experiment, we only used the *space* condition. In the second experiment, we were interested in the difference between the behavior of a person in an offender and a non-offender situation concerning self- and object-adaptors. To examine these behaviors, we used the *zoom* condition.

The results of this study indicate that offenders indeed show a different nonverbal behavior concerning these behaviors of interest. In the first experiment, offenders showed a different *moving pattern* than non-offenders: (1) Offenders in the authentic CCTV recordings and in the mock crime recordings were rated similar concerning *moving patterns*, (2) all offenders and non-offenders were rated to show *moving patterns* that deviate from *moving patterns* of the bystanders' crowd, (3) the strongest deviation from the bystanders' crowd was reached in the control search situation. *Communication behaviors* were not expressed differently in all three conditions (search, mock crime and real crime), but in all three conditions observed people's use of *communication behavior* deviated from the use of the bystanders' crowd. In the second experiment, offenders expressed more *object-adaptors* and less *self-adaptors* during the build-up phase of the mock criminal act compared to the search of a non-existing person.

2 "Who's the thief?" The influence of knowledge and experience on early detection of criminal intentions

Corinne I. Frey[1], Olive E. Wetter[2], and Franziska Hofer[2]

[1] University of Zurich, Cognitive Psychology Unit

[2] Kantonspolizei Zürich (Zurich State Police), Research & Development

Submission status

Submitted for publication.

Authors' contributions

CIF: Review of the literature, development of the research question and goals in close collaboration with FH, design of the study, development of the design of the test, selection of the recordings, preparing of the recordings, data collection, data analysis and interpretation, writing of the manuscript

OEW: Discussion of data analyses and interpretation, revising the manuscript

FH: Supervision and discussion of all of CIF's contributions, revising the manuscript

Acknowledgment

This research was funded by a grant from the Swiss federal office of civil aviation and the Kantonspolizei Zürich (Zurich State Police). The authors thank Markus Ruh for programming the computer-based test and the two criminal investigators who helped with choosing the stimuli. Special thank goes to the police officers who were always cooperative and supportive with insight into their daily work and experience. The authors would also like to thank Prof. Dr. Klaus Oberauer for revising the manuscript.

Abstract

The present study investigates if people are able to recognize thieves based on their nonverbal behavior prior to committing the crime. To this end, we implemented authentic CCTV footage from thefts committed at an international airport into a computer-based test. Five groups of participants (students, police recruits, inexperienced police officers, experienced police officers, criminal investigators) were studied. The results show that laypeople already performed above chance in the early recognition of thieves. However, they were outperformed by inexperienced police officers, who had just finished their basic training. Criminal investigators performed best. It follows that criminals display nonverbal behavior that can be used by observers for early recognition of criminal intentions. In addition, early recognition seems to benefit from knowledge about the criminals' *modi operandi*, which renders early recognition teachable and trainable. Further, all participants were biased towards innocence in their response tendency, but this bias was less pronounced in police officers than in students. These findings are discussed in relation to the in lie detection research well documented truth-bias and investigator-bias.

Keywords:

Behavior detection; early recognition of criminal intentions; interpretation of nonverbal behavior; truth-bias; investigator-bias

2.1 Introduction

The prevention of criminal activities is one of the core tasks of law enforcement in order to ensure public security. Preventative measures can involve the public, for example through information campaigns and educational work, or they can be carried out by authorities themselves, as in prevention and counteraction of criminal activities by early recognition of offenders. The earlier law enforcement personnel can detect criminal intentions on patrol or during an investigation, the more time for reaction and intervention remains, thus potentially increasing the chances of success. For example, if law enforcement personnel can identify thieves before the actual theft is committed, they can observe the action more clearly, record evidence while catching the thieves in the act, and ensure that the stolen goods do not disappear but can be handed back to the victims.

In the context of terrorist attacks, which have evolved into one of the biggest threats to society during the last decades, the prevention of criminal activities appears to be even more important. In that case, the observation and analysis of passenger behavior is a security measure that is becoming increasingly popular at airports in order to facilitate an early detection of signs of criminal intentions. In practice, several countries including pioneers like Israel and the U.S., as well as an increasing number of European countries, implemented programs that aim at improving early detection of signs of criminal intentions (Frank et al., 2009; U.S. Government Accountability Office, 2010). Although these programs are based on different concepts and work with different methods, they all train security personnel in the analysis of people's behavior for signs of criminal intentions, primarily to detect terroristic intentions. For example, the U.S.' program *SPOT* (*S*creening *P*assengers by *O*bservation *T*echnique) teaches a checklist of suspicious behaviors (especially nonverbal ones), which is then used by behavior detection officers for screening people and situations. Depending on the kind and the intensity of the observed behavior, law enforcement officers will take over

and conduct further investigation of the person in question (U.S. Government Accountability Office, 2010). At least three conditions must be met so that such prevention programs at airports or in general policing can be effective: First, offenders must display significant behaviors that relate to their intentions, and that can thus arouse suspicion in observers. Second, observers must have the ability to infer criminal intentions based on other people's (mostly nonverbal) behavior. Third, as it can be assumed that not everybody is already proficient by nature, the detection of criminal intentions through the analysis of the offenders' behavior must be teachable and trainable.

The correct interpretation of other people's behavior is an important ability for daily human life (Asch, 1946; Johansson, 1973; or Baldwin & Baird, 2001, for review). For instance, if a car driver decelerates and waves a hand at a pedestrian who wants to cross a road without traffic lights, the pedestrian can conclude from the driver's behavior that the driver intends to stop and give way. Without being able to infer other people's intentions based on their behavior, such nonverbal communication would not be possible. Thus, at least if nonverbal signs are used which deliberately aim at communicating intentions, we seem to be able to infer other people's intentions. Apart from such obvious nonverbal cues, other physical signs can serve as cues for inferring hypotheses about internal processes of others as well (Gray, 1988; J. Burgoon et al., 2005; Vrij, 2008; Eachus et al., 2013). For instance, trembling or perspiration can be read as signs for an increased stress level. Therefore, such physical signs can also provide information about others even though the counterpart does not consciously intend to provide this information (e.g., Zuckerman, DePaulo, & Rosenthal, 1981).

Such signs often appear as a byproduct of the physical reaction to a stressor. Imagine a thief being interrogated by the police. Whereas the primary intention in such a situation may be to deceive the interrogator, the interrogated person tries to *hide* this intention, for example

by trying to behave as naturally as possible (Ekman, Friesen, & O'Sullivan, 1988). The fear of this primary intention being discovered by the interrogator can lead to increased stress symptoms that cannot be (fully) controlled deliberately, since they are an autonomous reaction of the vegetative nervous system (Greene et al., 1985; Andreassi, 1995). Therefore, under such stressful circumstances it might be more difficult to control nonverbal behavior than verbal behavior (DePaulo, 1992; Ekman, 1993). For example, so-called micro-expressions (i.e., muscle action in the face due to strongly felt emotions) are such an involuntary sign which may reveal valuable information about deception (Ekman & Friesen, 1969a; Ekman, 2009). Hence, the interrogator might be able to read these physical (nonverbal) signs as hints that the person may actually be lying. Ultimately, the intention of deception may be revealed.

Several studies examined whether performance in the detection of lies is better than chance and, in particular, whether different groups of people with different amounts of expertise (e.g., laypeople, police officers, or expert interviewers) differ in performance (see Vrij, 2008, for review). This research revealed that the performance even of the most professional lie catchers is modest (DePaulo & Pfeifer, 1986; Vrij & Graham, 1997; Meissner & Kassin, 2002; Hartwig et al., 2004; Bond & DePaulo, 2006), suggesting that the human ability to detect deception in an interrogation setting is generally rather poor. Kassin, Meissner, and Norwick (2005) showed that lie detection performance is slightly better if observers have access to the audio channel only, thus suggesting that nonverbal behavior or physical signs might even distract from the correct interpretation. In sum, the current scientific literature on detection of lies neither consistently supports the assumption that nonverbal behavior or physical signs represent reliable sources of information for detecting criminal intentions, nor that professional experience leads to improvements in doing so.

In contrast to those seemingly discouraging results, Troscianko et al. (2004) found that people nevertheless seem to be able to predict aggressive behavior based on nonverbal cues, at least for antisocial unlawful acts like assaults or vandalism: They found that CCTV-operators as well as laypeople are able to predict antisocial or criminal, aggressive behavior in CCTV recordings. Contrary to situations that are used as examples in teaching the detection of lies, and contrary to situations in which an offender tries to act as inconspicuously as possible, the offenders in this study did not try to hide their intentions. Therefore, the task was not to detect a *hidden* criminal intention, but rather an intention of aggressive or antisocial behavior. From an evolutionary perspective, recognizing aggressive behavior as early as possible is important for one's own survival. This may explain why laypeople are as good as CCTV-operators in that particular task. Grant and Williams (2011) could not replicate these findings: Their participants only performed at chance level. Nevertheless, the idea survives that in some situations people indeed are able to predict unlawful, aggressive acts based on nonverbal behavior.

Much like liars in interrogation settings, smugglers (Vrij, 2008), thieves, and most terrorists try to hide their criminal intention as long as possible, i.e., they try to deceive. However, lie detection differs from detection of deception prior to committing a criminal act in one main aspect: The typical build-up phase that culminates in a criminal act as for example a theft, during which the offender tries to act as naturally as possible, is missing in situations like interrogations. Furthermore, research showed that during interrogations, both liars and truth-tellers alter their behavior strategically: liars do so in order to appear truthful, whereas truth-tellers do so in order to make sure the interrogator does not mistake them for liars (DePaulo et al., 2003; Strömwall et al., 2006). Regardless of their experience, it may therefore be difficult for interrogators to distinguish these two cases. In contrast, the task of distinguishing individuals with criminal intentions from innocent bystanders in a public space does not contain this problem. Thus, both the offender's as well as the observer's "task" differ

between lie detection and the detection of criminal intention prior to committing a criminal act.

The build-up phase prior to a criminal act can be illustrated by the example of Richard Reid, who attempted to detonate a bomb inside the cabin during a flight in December 2001 (Younge & Campbell, 2003; Frank et al., 2009). Reid had to enter the airport, pass the security screening, and enter the plane without being detected. His criminal intentions must have been present days before already. Thus, he had to hide his intentions for a long time. He presumably tried to act as naturally as possible in order not to be detected by the security personnel. In a similar way, criminals (e.g., thieves or drug dealers) have to hide their intentions as long as possible in order to be "successful." Thus, we accept the premise that smugglers, thieves, and most terrorists try to act as inconspicuously as possible before committing the crime, which may cause them to display similar nonverbal behavioral patterns.

There is only very little research examining the detection of *deceptive* offenders during the build-up phase. A small field study showed that a professional bodyguard indeed could detect mock terrorists based on their nonverbal behavior while they tried to hide their intention of attacking a public person (Heubrock et al., 2009a). Using CCTV footage of real thefts at a large international airport, we are the first ones to investigate 1) whether it is possible to detect *real hidden* criminal intentions during the build-up phase, as well as 2) the influence of different levels of professional expertise on detection performance.

Research in radiology (Nodine et al., 1999) showed an effect of expertise on the detection of cancerous tissue. In a similar way, studies in airport security screening found effects of training and expertise on the detection of threats inside passenger bags (McCarley, Kramer, Wickens, Vidoni, & Boot, 2004; Schwaninger & Hofer, 2004; Hardmeier, Hofer, & Schwaninger, 2006; Schwaninger, Hofer, & Wetter, 2007). These visual search tasks can be

seen as similar to the search for suspicious people in video footage because these tasks rely on the same basic cognitive mechanism: In order to be able to successfully perform these tasks, observers have to know exactly what they are looking for, i.e., they have to have a mental representation of the search target in visual memory. In the case of detecting criminal intentions prior to a criminal act, such a search target might be a particular nonverbal behavioral sign or a cluster of such signs (Frank et al., 2009). Thus, people with experience in that domain might perform better simply because they know what to look for. Nevertheless, neither Troscianko et al. (2004) nor Grant and Williams (2011) found a difference in detection performance between CCTV-operators and laypeople in the detection of antisocial or aggressive behavior. However, it might be the case that CCTV-operators do not have the same degree of systematic training in their specific task as compared to radiologists or X-ray screeners who are highly trained within their domain.

Much like radiologists or X-ray screeners, police officers (POs) gain very detailed knowledge about criminal activities during their theoretical training and with professional experience. They learn which *modi operandi* offenders use and which suspicious signs they display. Generally speaking, we expect POs to perform better than laypeople in detecting thieves before they commit the crime. We expect this effect, which is due to increased specific knowledge about crimes, to already be present for POs who have finished their basic training, but not for recruits who only just begin police training (Hypothesis 1).

Research on expertise by Ericsson and Lehmann (1996) and Ericsson (2008) emphasizes the importance of extensive daily experience and hard work in order to reach high levels of performance in that domain. Police work is very diverse in nature. Criminal investigators have the highest degree of specialization in and in-depth daily experience with detecting criminal behavior. Therefore, we expect criminal investigators to outperform other POs as well as laypeople in detecting thieves before they commit the crime (Hypothesis 2).

An effect of work expertise on the prediction of aggressive criminal behavior has neither been found by Troscianko et al. (2004), nor by Grant and Williams (2011). Besides the above mentioned possible lack of enough systematic training of CCTV-operators in detecting criminal intentions, maybe these studies could not reveal an effect of expertise simply because expertise was measured with detection performance *at the very end of the video footage*. It may well be the case that expertise manifests itself in early detection, probably even during the build-up phase. We expect that POs detect criminal intentions earlier than laypeople (Hypothesis 3).

Research on lie detection has identified two biases: the *truth-bias* (e.g., McCornack & Parks, 1986) and the *investigator-bias* (e.g., Meissner & Kassin, 2002). People's responses tend to be biased towards judging messages more often as truths than lies, revealing the so-called *truth-bias* (Anderson, Ansfield, & DePaulo, 1999). Anderson et al. (1999) argue that the truth-bias might be the result of everyday life experience, where people are not often confronted with lies. Meissner and Kassin (2002) have revealed the *investigator bias:* POs are biased more towards deception than laypeople. In other words, POs have the tendency to judge messages as lies more often than laypeople. It may be assumed that POs' general suspicion might be raised more easily compared to laypeople because they encounter deception on a daily basis. The truth-bias and the investigator-bias raise the question of who is actually biased: is it the general population, the POs, or everybody? Our study aims at reconciling these views by studying the response tendencies of both groups in the domain of detecting hidden criminal intentions and using the same task for laypeople and POs. For the moment, we expect that POs are more suspicious than laypeople (Hypothesis 4).

2.2 Method

2.2.1 Participants

A total of *N* = 315 participants were tested. They belonged to five different groups regarding their knowledge and experience with crime. Below, the five different groups and their characteristics are explained. Detailed information on age and gender for each group is given in *Table 1*. All participants received immediate feedback about their results at the end of the test. They were informed of the possibility to withdraw their consent at any time.

Table 1
Experience, age, and gender for the five groups of participants.

Group	Experience in police work in years (mean, min-max)	Age in years (mean, min-max)	Gender (male; female)
Students of psychology (*n*= 50)	none	24.72 (20-36)	25; 25
Police recruits (*n* = 40)	none	25.48 (20-34)	31; 9
Inexperienced POs (*n* = 129)	0.45 (0.16-0.75)	26.10 (21-37)	90; 39
Experienced POs (*n* = 51)	13.91 (4-32)	38.80 (25-56)	46; 5
Criminal investigators (*n* = 45)	14.22 (4-31)	38.62 (28-54)	41; 4

Fifty students of psychology served as a first group of laypeople. In order to exclude possible gender effects, we fully balanced the gender ratio in this group. They were either paid according to the usual conditions at the University of Zurich (i.e., CHF 15 per hour) or received course credit. As a second group of laypeople, we tested 40 police recruits who had just started their basic police training. These two groups did not differ regarding their experience or knowledge about criminals and their *modi operandi*. However, it can be expected that the police recruits are highly motivated for this task. Comparisons between these two groups of laypeople will thus later allow us to discover potential motivational effects as well as self-selection effects, meaning that people might join law enforcement, because they believe that they are really good at detecting criminals.

Next, we included professional POs of different stages in their career in our quasi-experimental design: We examined 129 inexperienced POs who had just finished the first year of their training, which consists of basic theoretical training and a few months of practical experience only, 51 POs who never worked as criminal investigators but who are considered as generally experienced, as well as 45 criminal investigators. This last group consisted of people who were all active criminal investigators in the respective branch of Zurich State Police at the time the study was conducted, or people who had worked there before during their career. All POs took the test during their paid working hours.

2.2.2 Materials

A computer-based test was developed that employs authentic CCTV footage showing baggage thieves shortly before they commit the crime. We decided to use authentic CCTV footage of thefts for three reasons: First, it is hypothesized that thieves try to hide their intentions as long as possible in order to avoid being detected by bystanders or the victims themselves. Second, as POs participated in the study, we wanted to make sure that the used video footage shows crimes they are familiar with. As thefts occur frequently, most POs encounter them during their career. Third, it is important to test detection performance with real crimes because it is not yet clear whether mock criminals show similar behavior to real criminals.

Twelve CCTV recordings from a large international airport were used as stimuli. Each recording contained a real theft committed by one to three baggage thieves in either the check-in or the shopping area. In all recordings together, there were 23 different thieves. The recordings were selected based on the following criteria:

a) The build-up phase of the theft lasts at least 50 s.

b) Three or more possible distractors are present. This means that there are three or more other people, who remain visible for a similar amount of time as the thieves. They act a little different from the surrounding crowd (e.g., standing still while all other people are moving).

The first author and two POs from the criminal investigation branch of Zurich State Police, who did not take part in the study, selected the video footage used in this study. Each recording contained between twenty and over a hundred people in total. As the recordings were shown to the participants in original CCTV quality, no faces could be identified.

The twelve recordings lasted from 53 s to 144 s with an average of 104 s. Each recording was cut into three sequences. The third sequence stopped just before the first movement of the offender(s) towards the stolen goods or towards the victim. The decision about where to cut the first and the second sequence was again taken by the first author together with two experienced criminal investigators. This decision depended on the thieves' actions as there had to be at least some minimum action (e.g., body movement) present within the first sequence. Furthermore, we tried to keep the duration of the three sequences as similar as possible across all recordings: Sequence 1 lasted from 10 s to 52 s (M = 29 s), sequence 2 lasted from 10 s to 80 s (M = 37s), and sequence 3 lasted from 11 s to 90 s (M = 37 s).

2.2.3 Procedure

All POs were tested at the beginning of a two-day training course on the detection of suspicious behavior. Groups of 35 to 39 participants took the test together in a room with desktop computers. All participants received basic information on the relevance of the topic for security in general, as well as some basic information on our study. Exactly the same information was given to the two laypeople groups, who were tested in groups of one to five (students of psychology) and 40 people (police recruits), respectively.

The computer-based test first displayed an instructional text followed by an example recording including a theft. Participants were informed about the existence of one to three thieves in each movie, but they did not know the exact number of thieves in each of the three sequences. After the example movie, participants had the possibility to ask questions or to continue directly by pressing the space bar. The twelve recordings were presented in random order. Each sequence ended with a still frame. At this point the participants had to mark the suspected person(s) by clicking the mouse on the head(s) of the suspect(s). By pressing the space bar, participants started the next sequence. If participants decided that no offender was visible in a sequence, they omitted the mouse click and just pressed the space bar to start the next sequence. After each sequence and after each answer, participants were allowed to take a short break if they wanted to. At the end of the test, participants received feedback on the number of correctly identified thieves per sequence.

2.3 Results

In our analyses, we employed measures from Signal Detection Theory (Green & Swets, 1974), which can be seen as a standard in research on detection performance. Based on this theory, there are four possible combinations of a participant's answer and stimulus presence: In our case, a correctly identified thief is defined as a *hit*, a non-thief that did not arouse suspicion is defined as a *correct rejection*, a wrongly suspected person is defined as a *false alarm*, and a missed thief is defined as a *miss*. For methodological reasons, the number of total clicks within each sequence was limited to three, i.e., in each sequence, a maximum of three false alarms and of one to three hits was possible. As a consequence, 108 false alarms were possible overall. Participants could score 20 hits in the first sequences of all recordings, 22 hits in the second sequences, and 23 hits in the third sequences. As a consequence, 65 hits were possible overall. The measures d' and C were used to analyze detection performance and response tendency, respectively. As they are calculated based on percentages of hits and false

alarms, the different absolute numbers of possible hits and false alarms are accounted for. Because sphericity was never given in our data, we corrected the degrees of freedom according to Greenhouse-Geisser in all subsequent analyses. In all multiple comparisons, we adjusted the *p*-value according to Bonferroni.

2.3.1 Hits and false Alarms

As a first step, we plotted the overall hit rate (pHit) and the overall false alarms rate (pFA) for each participant (*Figure 1*).

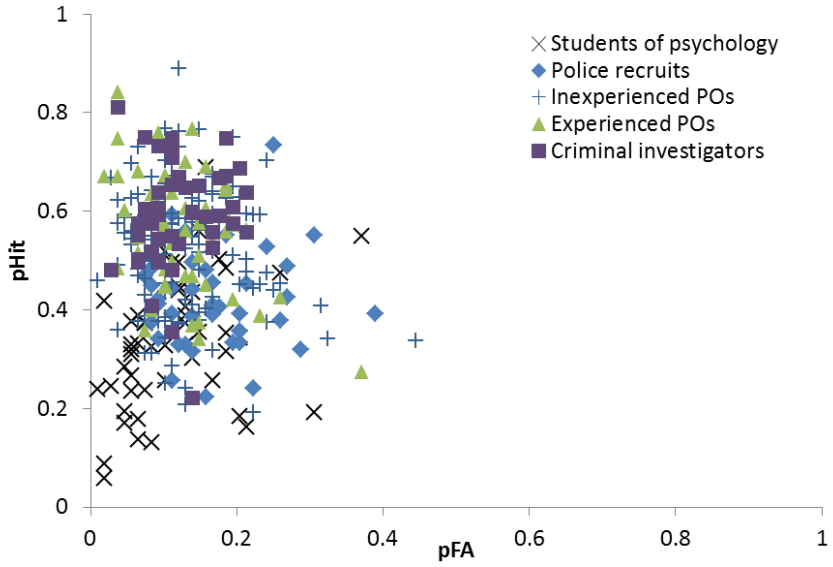

Figure 1. Overall hit and false alarm rates for all five groups.

All participants had a rather low pFA, whereas their pHit spread over almost the whole possible range. We then compared all five groups in all three sequences with respect to their pHit as well as their pFA. In order to do so, we conducted two mixed analyses of variance (ANOVA) with group (students of psychology, police recruits, inexperienced POs, experienced POs, and criminal investigators) as between-subjects factor and sequence (1, 2, 3) as within-subjects factor (*Figure 2*).

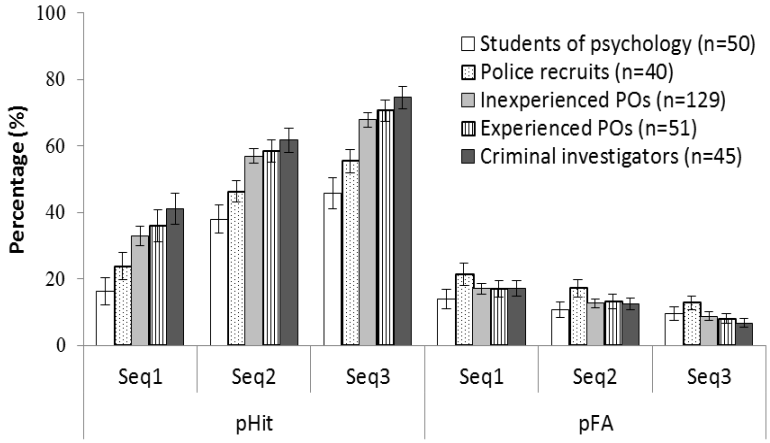

Figure 2. Mean hit and false alarm rates for all five groups in all three sequences. Error bars represent confidence intervals (95%) for between-subjects comparisons.

The ANOVA for pHit shows a significant main effect of group with $F(4, 310) = 38.21$, $p < .001$, $\eta^2 = .33$, and a significant main effect of sequence with $F(1.53, 473.13) = 931.40$, $p < .001$, $\eta^2 = .75$. There is no significant interaction between group and sequence $F(6.11, 473.13) = 1.39$, $p = .217$, $\eta^2 = .02$, indicating that the increase in pHit with increasing sequence is similar for all groups in all sequences. Post hoc pairwise comparisons between the five groups are displayed in *Table 2*. All groups of POs and the police recruits differ significantly from the students of psychology in pHit (all PO groups: $p < .001$, police recruits: $p = .010$). Police recruits differ significantly from all PO groups in pHit (all comparisons at least: $p < .05$). The criminal investigators differ significantly from inexperienced POs in pHit ($p = .019$), but not from experienced POs ($p = 1.000$).

Table 2
Pairwise comparisons of pHit between students of psychology, police recruits, inexperienced POs, experienced POs, and criminal investigators (N = 315).

	Police recruits	Inexperienced POs	Experienced POs	Criminal investigators
Students of psychology	$p = .010$	$p < .001$	$p < .001$	$p < 0.001$
Police recruits		$p < .001$	$p < .001$	$p < 0.001$
Inexperienced POs			$p = 1.00$	$p = .019$
Experienced POs				$p = 1.000$

The ANOVA for pFA shows a significant main effect of group with $F(4, 310) = 4.88$, $p < .001$, $\eta^2 = .06$, and a significant main effect of sequence with $F(1.67, 520.04) = 143.94$, $p < .001$, $\eta^2 = .32$. Contrary to pHit, pFA decreases over all three sequences as can be seen in *Figure 2*. There is a significant interaction between group and sequence $F(6.71, 520.04) = 2.39$, $p = .022$, $\eta^2 = .03$. The detailed values of post hoc pairwise comparisons between the five groups are displayed in *Table 3*. All groups differ significantly from the police recruits in their pFA (experienced POs: $p < .05$; all other groups: $p < .01$).

Table 3
Pairwise comparisons of pFA between students of psychology, police recruits, inexperienced POs, experienced POs, and criminal investigators (N = 315).

	Police recruits	Inexperienced POs	Experienced POs	Criminal investigators
Students of psychology	$p < 0.001$	$p = 1.000$	$p = 1.000$	$p = 1.000$
Police recruits		$p = .004$	$p = .021$	$p < 0.001$
Inexperienced POs			$p = 1.000$	$p = 1.000$
Experienced POs				$p = 1.000$

2.3.2 Detection performance, d'

Based on pHit and pFA, detection performance d' was calculated (*Figure 3)*. We then compared d' of all five groups in all three sequences with a mixed analysis of variance (ANOVA). Group served as between-subjects factor and sequence served as within-subjects factor. The results are depicted in *Figure 3*. We also conducted pairwise comparisons of all five groups.

There is a significant main effect of group with $F(4, 310) = 22.10$, $p < .001$, $\eta^2 = .22$, and a significant main effect of sequence with $F(1.84, 568.75) = 860.74$, $p < .001$, $\eta^2 = .74$. Thus, the more of the recording the participants had seen, the more information they could gather. Furthermore, the ANOVA shows a significant interaction between sequence and group, with $F(7.34, 568.75) = 3.40$, $p = .001$, $\eta^2 = .04$, indicating that at least one group

benefits more from the additional information provided in higher sequences than other groups. *Figure 3* shows that laypeople's detection was at chance level in sequence 1, whereas inexperienced POs performed above chance level already. In sequence 2, the laypeople's detection performances increased rather strongly while the other groups' performance increased less. This finding is confirmed by testing detection performance against chance level using one-sample t-tests: In sequence 1, students of psychology ($t(49) = 0.10$, $p = .919$) and police recruits ($t(39) = 1.16$, $p = .252$) did not perform above chance level, whereas inexperienced POs showed performance above chance level ($t(128) = 9.50$, $p < .001$). In sequence 2, the performance of the students of psychology ($t(49) = 11.82$, $p < .001$) and the police recruits ($t(39) = 11.64$, $p < .001$) increased above chance level too.

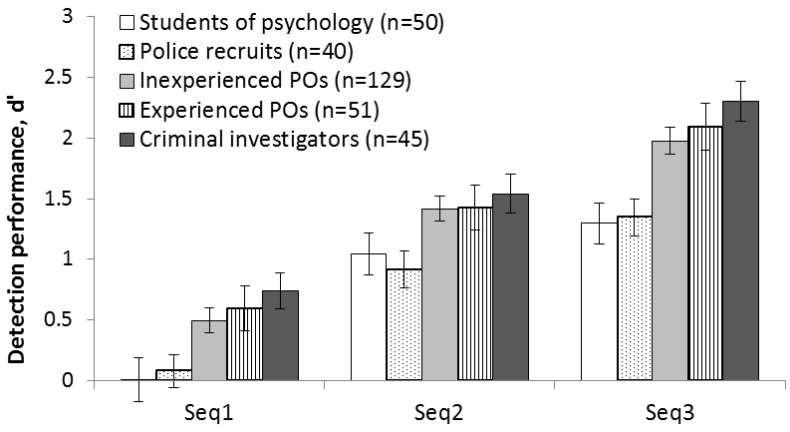

Figure 3. Detection performance (d') of the five groups in the three sequences (means). Error bars represent confidence intervals (95%) for between-subjects comparisons.

The pairwise comparisons between the different groups, as given in *Table 4,* indicate no significant difference between students of psychology and police recruits ($p = 1.000$; two-tailed). The performance of these two laypeople groups differs significantly from all PO groups, though (all comparisons $p < .001$; one-tailed). Criminal investigators outperformed inexperienced POs ($p < .05$; one-tailed), but not experienced POs ($p = .500$; one-tailed).

Table 4
Pairwise comparisons of detection performance d' between groups (N = 315).

	Police recruits	Inexperienced POs	Experienced POs	Criminal investigators
Students of psychology	$p = 1.000$	$p < .001$ [a]	$p < .001$ [a]	$p < .001$ [a]
Police recruits		$p < .001$ [a]	$p < .001$ [a]	$p < .001$ [a]
Inexperienced POs			$p = .500$ [a]	$p = .048$ [a]
Experienced POs				$p = .500$ [a]

[a] *One-tailed significance based on hypotheses*

2.3.3 Response tendency, C

Data on response tendency C are given in *Figure 4*. We conducted a mixed ANOVA with group (students of psychology, police recruits, inexperienced POs, experienced POs, and criminal investigators) as between-subjects factor and sequence (1, 2, 3) as within-subjects factor. This analysis shows a highly significant effect of group, $F(4, 310) = 1699.22$, $p < .001$, $\eta^2 = .85$. There is also a highly significant effect of sequence, $F(1.43, 442.49) = 121.23$, $p < .001$, $\eta^2 = .28$. *Figure 4* reveals that, the higher the sequence number, the more neutral the participants' response tendency. The interaction between sequence and group reaches marginal significance with $p = .056$, which hints at a possible change in response tendency over the three sequences for at least one group.

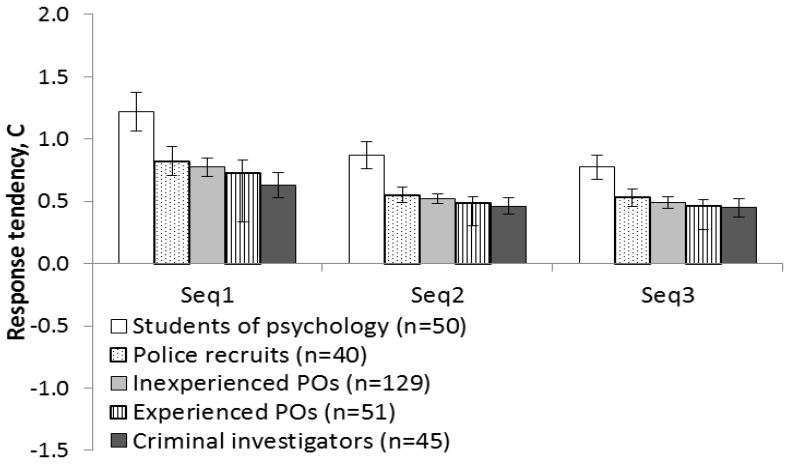

Figure 4. Response tendency (C) of the five groups in the three sequences (means). Error bars represent confidence intervals (95%) for between-subjects comparisons.

Furthermore, *Figure 4* reveals that POs and police recruits are closer to a neutral criterion than the students of psychology. The pairwise comparisons confirm this difference: The response tendency of the students of psychology is significantly more biased towards innocence than the response tendency of all other groups (all comparisons $p < .001$). Still, all groups are significantly biased towards innocence. One-sample t-tests for sequence 3, that tested the groups' criteria against zero, all reveal significant results: police recruits, $t(39) = 14.48$, $p < .001$; inexperienced POs, $t(128) = 21.53$, $p < .001$; experienced POs, ($t(50) = 17.52$, $p < .001$; criminal investigators, $t(44) = 11.57$, $p < .001$.

2.4 Discussion

This pioneering study investigates whether criminals, who try to hide their criminal intentions, can be detected by the means of behavioral analysis prior to committing the crime. We examined the influence of theoretical knowledge and professional expertise of POs on detection performance and response tendency in behavioral analysis. To this end, we tested POs with different levels of experience, as well as laypeople, by the means of a computer-based test that featured authentic CCTV footage showing thieves shortly before they successfully managed to steal.

The results show that thieves can indeed be detected prior to committing the crime based on their behavior. All groups of participants were able to detect the thieves significantly above chance level shortly before committing the theft (namely in sequences 2 and 3). Contrary to criminals that plan an aggressive assault, thieves attempt to hide their intention by trying to act as natural or unsuspicious as possible. Hence, our results extend the results of Troscianko et al., (2004) who showed that aggressive criminal intentions can be detected based on nonverbal signs: We show that the same is true for *deceptive* criminal intentions. Furthermore, our study shows that it is possible to identify criminal intentions of one or more offenders even in a rather large crowd of people. This is not yet possible for technical solutions such as automated detection software like Cogito, which can only be used in interrogation settings (Arsíc, Schuller, & Rigoll, 2007; Eachus et al., 2013).

Contrary to previous findings in behavior detection (Troscianko et al., 2004) and lie detection (for reviews see Kraut, 1980; Vrij, 2008), we found a significantly higher detection performance in groups with more knowledge and experience. Already inexperienced POs performed significantly better than students of psychology as well as police recruits. Thus, Hypothesis 1 is confirmed. Our results suggest that POs' theoretical training not only improves detection performance in behavior analysis to a remarkable degree, but that it also allows them to perform at the same level as experienced POs. This improvement in detection performance manifests itself particularly in a very early detection during the build-up phase, i.e., in contrast to laypeople, inexperienced POs are able to recognize thieves already relatively long before they commit the crime.

This finding indicates the importance of the point in time when performance is measured. In the studies by Troscianko et al. (2004) and Grant and Williams (2011), participants answered at the end of the build-up phase only. This could have masked a possible difference between CCTV-operators' and laypeople's performance. Another possible

explanation for the different results in our study concerning expertise might be POs' specific training and knowledge about criminal acts. This degree of specificity and depth is likely not to be reached in CCTV-operator training, which may explain why they did not perform better than laypeople in the aforementioned studies. This result is of high practical importance because it shows that theoretical knowledge, which is acquired during basic police training, already significantly improves detection performance. This implies for police practice that basic police training indeed provides POs with an advantage in order to allow for early intervention at crime scenes.

Hypothesis 2 was only partially confirmed: Criminal investigators outperformed all groups of participants except experienced POs. Thus, experienced POs perform somewhere in between inexperienced POs and criminal investigators. Experienced POs have been on the job approximately as long as criminal investigators, but, contrary to criminal investigators, were not able to outperform inexperienced POs. Thus, work experience in policing seems to be important in order to successfully interpret behavior. It remains unclear, though, how specific this work experience in policing has to be in order to be helpful in this particular task, because police work is very diverse per se. Ericsson (2008) proposed that not only many years of practice but also specific types of training and deliberate efforts are needed in order to reach an outstanding performance.

In our study, the criminal investigators represent the group which is most specialized in interpreting behavior during the build-up phase since this is a main task of their daily work. Their pattern of pHit and pFA (*Figure 1*) is rather homogeneous and located in the top half of the hit rate scale. The pattern of the experienced POs appears more scattered. In other words, the group of criminal investigators answered more homogenously on a rather high level while experienced POs showed larger inter-individual differences in their response patterns. This could be an indicator for higher specialization although there was no significant difference in

mean performance between these groups. The groups of experienced POs and criminal investigators had a large range of work experience (4 to 32, 4 to 31 years respectively). It appears possible that this could have masked group differences in performance.

We furthermore found that laypeople performed significantly above chance level (sequences 2 and 3), which is not in line with earlier findings from research on lie detection (DePaulo & Pfeifer, 1986; Vrij & Graham, 1997; Vrij & Mann, 2001; for review see Vrij, 2008). It seems that detecting deceit in interrogation settings is more difficult than detecting deceit during the build-up phase of a criminal act. One possible explanation for this might be that, during an interrogation, all people (liars and truth-tellers alike) try to convince the interviewer that they are telling the truth. Strömwall et al. (2006) found that truth-tellers as well as liars mention similar strategies in verbal and nonverbal behavior, such as *not making any excess movements* and *maintaining eye contact*, in order to persuade the interviewer of their innocence. As a consequence, it becomes rather difficult to decide which statements are actually true. The idea of both, truth-tellers and liars, altering their behavior during interrogations, and the resulting absence of a reliable guideline for behavioral cues underlying truthful statements, might explain the rather poor performance in lie detection of law enforcement and students.

The fact that observers in our study had the opportunity to directly compare possible suspects to other people who behaved normally might also provide a reason, why law enforcement performed rather well in our study as compared to Strömwall et al. (2006). In an interrogation setting as used by Strömwall et al. (2006), the observer interrogates a person without being able to compare the behavior directly to innocent people's natural behavior. In our setting, offenders probably used certain strategies in order not to get caught, i.e., they tried to behave as naturally as possible. Paradoxically, however, trying to act naturally might in fact render them highly suspicious because all the other people do not follow any strategies but

rather concentrate on catching their plane. This illustrates that innocent people display a reference behavior that can be taken as "baseline", i.e., behavior that is not influenced by any goal of making a certain impression on an observer.

In order to detect deceptive behavior regarding lies, it seems helpful to observe clusters of behaviors instead of just one behavior or behavioral pattern in isolation (Vrij, 2008). This statement is supported by own personal conversations with experienced criminal investigators. They reported that they perceive the scene as a whole and try to detect clusters of behavior that differ from the baseline. Findings of Gestalt psychology about figure-ground phenomena (Goldstein, 2007; Heubrock, Immerini, Mengeringhausen, & Palkies, 2009) support this approach. The visual system identifies objects using their outlines to discriminate between different objects. This is even possible if the outlines are not clearly visible; the visual system tries to make sense of the visual input, whereas prior experience influences what is perceived (Goldstein, 2007). Applied to the detection of suspicious behavior, the situation can be described as follows: observers are watching a crowd of people. If the observers do not have any prior knowledge about how criminals behave in preparing their crime, they cannot be detected. Detection is impossible or unlikely, because the observers' visual system could not make sense of the information available and, thus, figure-ground segregation was impossible (Heubrock et al., 2009).

However, if the observers know about the *modus operandi* (e.g., how a theft is committed), they may recognize a pattern which matches stored information. In this way they are able to distinguish the criminal as the "figure" from the background. This could explain our finding that inexperienced POs already, shortly after having completed the theoretical training, are able to detect thieves earlier in the build-up phase than laypeople. Thus, hypothesis 3 is confirmed. Apparently, the theoretical training had provided them with information on how thieves act so that they were subsequently able to detect their behavior as

"figure". It may be that the representation strength of criminal behavior increases with training and/or professional experience. This might explain why criminal investigators performed better than inexperienced POs in our task. Our personal conversations with inexperienced POs and criminal investigators suggest that the criminal investigators' knowledge about the *modus operandi* of thieves is far more sophisticated than the inexperienced POs' knowledge.

The existence of the *investigator-bias* is well established in lie detection research (Meissner & Kassin, 2002). The *investigator-bias* describes a response tendency towards suspecting lies rather than the truth. The results of Meissner and Kassin's (2002) meta-analysis showed that responses of trained or experienced people are particularly biased towards lies. Furthermore, we know from Masip, Alonso, Garrido, and Antón, (2005) that experienced POs show higher scores on a generalized communicative suspicion scale, which measures the general suspiciousness of people, than undergraduates and even police recruits. Based on these findings, we had expected our POs to be more biased towards suspicion than laypeople as well (Hypothesis 4). The results disprove this assumption: Police recruits judge behavior already more often as suspicious as students of psychology do, even though they are still laypeople. Thus, the explanation offered by Anderson et al. (1999) for this phenomenon, namely that the exposition to lies in actual daily life is responsible for the shift in response tendency, seems to not hold true in our context of the detection of criminal intentions: police recruits, who are not more exposed to criminals than students of psychology, have shown the criterion shift as well. It appears thus likely that other factors such as the mindset or the motivation to find criminals contribute more to the criterion shift than target prevalence in daily life. Either Anderson et al.'s offered explanation is wrong, or different mechanisms are at play in the detection of lies versus criminal intentions.

The question of whether there is a bias towards innocence (similar to the truth-bias in lie detection) or a bias towards suspicion (similar to the investigator-bias in lie detection) in our data can easily be reconciled as follows: The question of who is biased in which way fully depends on the viewpoint one takes. From the students' viewpoint, POs and police recruits have a suspicion-bias, since their response tendency C has significantly lower values. From the POs' and the police recruits' viewpoint, the students have an innocence-bias since their response tendency C has significantly higher values. From a neutral and scientific viewpoint, all groups are biased towards innocence (truth-bias) as all criterion values are positive and differ significantly from zero and that this bias is less pronounced in POs as well as police recruits compared to students of psychology.

Our finding that detection performance improves with knowledge and expertise, while the response tendency seems to be rather influenced by one's mindset and motivation to find criminals, extends previous research in lie detection (Meissner & Kassin, 2002; Masip et al., 2005; see Vrij, 2008 for review) as well as behavior detection (Troscianko et al., 2004; Grant & Williams, 2011). In our view, the fact that there was a change in criterion already in the police recruits does not contradict the effect of target prevalence as described by Wolfe et al., (2007). We believe that target prevalence can serve as a reminder that stimuli exist and occur, which subsequently keeps awareness at a high level. Awareness might also be raised by mentally preparing (Gould, Flett, & Bean, 2009) for a task, though, which is exactly what police recruits do. Thus, in the case of the police recruits, the missing target prevalence in daily life may have been substituted by mental preparation about criminals. However, this increased awareness in police recruits did not have a positive influence on police recruits' performance per se. Detection performance improved only with theoretical knowledge and work experience. This leads to the conclusion that in this case training can indeed improve detection performance independent of response tendency.

Whenever different populations are compared it can be argued that one population is more inherently motivated to perform well in the given task (Biggs, Cain, Clark, Darling, & Mitroff, 2013). In our study, POs could have been more motivated to perform well than students of psychology because preventing crimes is one of their main tasks. However, police recruits' performance did not differ from the performance of students of psychology. The difference in response tendency, which could be an indicator for more eagerness to suspect people, did not affect detection performance. We thus believe that potential differences in motivation between the groups cannot explain the differences in performance we found.

With the limitation that we only used one type of crime, namely theft, the present study empirically shows that it is possible to detect people who try to deceive based on their behavior before they commit the crime at least when they are observed on CCTV. It remains to be proven that this is also the case in real life. The kind of information that leads to this detection, i.e., whether there are specific signs or whether it is a more general kind of information, such as differences in behavior from the surrounding people, clearly warrants further study. Troscianko et al. (2004), for example, replayed some of the recordings to their participants and asked them to indicate when their suspiciousness was triggered. Afterwards, they analyzed the recordings frame by frame and found some nonverbal behavior cues to trigger suspiciousness rather often (i.e., distinctive gaits and hand gestures). It still remains to be investigated more thoroughly whether such detectable behaviors can indeed be specified.

Altogether, our study gives important insights on the influence of knowledge and experience on people's abilities to perform a complex visual search task, namely detecting criminals within a crowd based on their nonverbal behavior. In particular, our study provides basic findings that are highly relevant to general policing as well as behavior detection programs that are implemented at many airports. Our study showed that, firstly, offenders trying to be unsuspicious (i.e. trying to *hide* their criminal intentions) show behavior that

triggers suspicion. Secondly, observers can indeed detect *hidden* criminal intentions during the build-up phase of a criminal act based on nonverbal information only. Thirdly, this detection ability can be improved by training and practice. Further research is needed in order to obtain more information on which contents of the theoretical training lead to this increase in performance.

3 What is suspicious when trying to be inconspicuous? Criminal intentions inferred from nonverbal behavioral cues.

Corinne I. Frey[1], Olive E. Wetter[2], and Franziska Hofer[2]

[1] University of Zurich, Cognitive Psychology Unit

[2] Kantonspolizei Zürich (Zurich State Police), Research & Development

Submission status

Submitted for publication.

Authors' contributions

CIF: Review of the literature, development of the research question and goals in close collaboration with FH, design of the study, organizing the film crew for own recordings, selection of the recordings, supervising the preparing of the recordings, programming the experiment with Tatool, data collection, data analyses and interpretation, writing of the manuscript

OEW: Discussion of data analyses and interpretation, revising the manuscript

FH: Supervision and discussion of all of CIF's contributions, organizing mock offenders for own recordings, revising the manuscript

Acknowledgment

This research was funded by a grant from the Swiss federal office of civil aviation and the Kantonspolizei Zürich (Zurich State Police). The authors thank the film crew consistent of several police officers for their work. The authors also like to thank all actors for taking part and allowance of the use of their recordings. Special thanks go to the police officers from the Kantonspolizei Zürich who were always cooperative and supportive with insight into their daily work and experience. The authors also thank Simona Seidmann for her assistance with data collection.

Abstract

The present study investigates whether nonverbal behavioral cues to hidden criminal intentions during the build-up phase of a criminal act exist. To this end, we created recordings of actors once in a search situation and once committing a mock crime (theft or bomb placing) in a public crowded area. For ecological validation, we used authentic CCTV footage from thefts committed at an international airport in Experiment 1. In this experiment, *moving patterns* in public space and *communication behavior* were analyzed according to the difference of the expressed behavior in search and (mock) crime situations as well as compared to a bystanders' baseline. Experiment 2 was conducted to examine *object-adaptors* and *self-adaptors* as the behaviors of interest while controlling for inter-personal differences. The results of this study indicate that nonverbal behavior of offenders differs from nonverbal behavior of non-offenders. However, this holds only under the conditions of a valid baseline and of judging not only *a* typical behavior but a cluster of nonverbal behaviors as behavioral indicators during the build-up phase of a criminal act.

Keywords:

Behavior detection; cues to hidden criminal intentions during the build-up phase; interpretation of nonverbal behavior; criminal intentions

3.1 Introduction

Crime prevention is one of the police's most important tasks. In order to perform well in this task, police officers need to be able to detect offenders before they commit a crime. As officers are usually not involved in conversations with offenders during the build-up phase of a crime, the detection most often relies on nonverbal behavioral cues as the only source of information. Crimes such as thefts, drug deals, and terrorist attacks, most often happen in public spaces. They can happen at train stations or airports with a large crowd of bystanders who are unaware of the crime being committed. Thus, in many cases, officers need to observe crowds of people in order to detect offenders before they commit the crime. During the build-up phase, most offenders try to *hide* their criminal intentions to be successful. For example, thieves enter a public space with the criminal intention to steal valuables from an innocent person. In order to be successful, they need to appear inconspicuously since neither the future victim nor bystanders or security personnel must infer their criminal intentions. The attacks on the World Trade Center on September 11, 2001, for example, included a successful build-up phase for every team of hijackers. Presumably, the intention to hijack these planes existed for a rather long time and the attack was very well planned (National Commission on Terrorist Attacks Upon the United States, 2004). The hijackers had to pass all security protocols on the airports successfully in order to fulfill their plan. It is known that some of these terrorists managed to hide their criminal intentions even though they were questioned by security personnel during the build-up phase (National Commission on Terrorist Attacks Upon the United States, 2004).

These attacks in 2001 contributed to shaping the public awareness of terrorism being one of the biggest threats to society. Thus, with the U.S. and Israel taking the lead, several countries developed intelligence programs which are based on behavior detection during the build-up phase (U.S. Government Accountability Office, 2010). For example, the behavior

detection program of the U.S., SPOT (*S*creening *P*assenger by *O*bservation *T*echnique), is based on a checklist with (mostly nonverbal) cues which should indicate criminal intentions. Based on this checklist, behavior detection officers screen airports and other vulnerable infrastructure for these cues of interest. Firstly, the observers establish the baseline values for these cues in the current situation. In other words, they judge how strongly or how frequently the specified behavioral cues are expressed in the observed crowd overall. People who show nonverbal behavioral cues that added up reach a certain threshold are selected for further investigation by the police (Frank et al., 2009; U.S. Government Accountability Office, 2010). As pointed out by Frey, Wetter, and Hofer (submitted), three premises have to be given for such programs to be effective: a) Offenders show nonverbal behavioral cues which indicate their criminal intentions and deviate from the baseline, b) Observers must be able to detect and interpret these cues correctly, and c) the detection and correct interpretation of such cues have to be teachable or trainable.

Contrary to the detection of lies (see Kraut, 1980; Vrij, 2008 for review), detection performance of criminal intentions is above chance level. Frey et al. (submitted) showed that people are indeed able to detect *deceptive* offenders during the build-up phase. This extends the results of Troscianko et al. (2004), who found that it is possible to infer aggressive criminal behavior based on nonverbal cues. Troscianko's participants were shown CCTV recordings of the build-up phase of either aggressive unlawful acts, lawful social interactions which were similar to the unlawful acts (matches), or different social interactions. They were then asked to predict the outcome of the situations, in which they performed above chance level. Frey's participants, however, were shown CCTV recordings of the build-up phase of real baggage thefts committed in a rather large crowd of people. The participants were given the opportunity to identify offenders at three different points in time during the build-up phase. Contrary to Troscianko et al. (2004), Frey's participants knew that a crime (theft) would happen. They had to correctly identify the thieves among the innocent bystanders as

early as possible. The results show that laypeople perform above chance level shortly before the theft was committed. Police officers with different levels of professional experience outperformed laypeople significantly on all three time points during the measured build-up phase. The study indicates that police officers possess valuable knowledge about offenders' behavior that seems to be helpful.

The idea of nonverbal behavioral cues helping to identify hidden criminal intentions related to deception is very well studied within the area of lie detection research (Zuckerman et al., 1981; DePaulo et al., 2003; see Vrij, 2008 for review). Laypeople and professionals alike strongly believe in a reliable correlation between diagnostic nonverbal behavioral cues and lying (Mann et al., 2004; Sporer & Schwandt, 2007), although research mostly contradicts these beliefs (Vrij, 2008). Zuckerman et al. (1981) proposed the following three important factors, which influence nonverbal behavior and, therefore, can serve as possible cues in order to identify deception: a) emotional reactions, b) cognitive effort, and c) attempted behavioral control.

Ekman and Friesen (1969b) and Ekman et al. (1987) showed that the facial expressions of basic emotions such as anger, happiness, fear, sadness, surprise, and disgust, are universal. In their work the authors also mention *display rules,* which are rules for managing facial expressions that are determined by culture and society. For example, as soon as an authorial person was placed in the same room as Japanese participants, they masked their felt emotions by smiling more (Ekman, 1998). Based on these results, it can be assumed that people might be able to control their nonverbal behavior, or at least their facial expressions, to some extent. More importantly, people tend to try to control their behavior under certain circumstances (J. K. Burgoon & Buller, 1994), and they try to display behavior that they believe is not associated with lies (Sporer & Schwandt, 2007).

According to Ekman (2009), the basic emotions underlying lies are guilt, fear, and delight. Fear, for example, could result in an increase of physical arousal and entail an increase of nonverbal behavior that indicates arousal. According to Vrij (2008), this could cause a higher frequency of cues such as eye blinks and/or self-adaptors (e.g. touching own hair, face). Nevertheless, negative emotions, such as fear and guilt, could also lead to withdrawal signs (Vrij, 2008). The same internal state could therefore lead to different nonverbal behavioral cues and this indicates the complexity of human behavior. In addition, many different internal states may underlie a hostile or criminal intention and these different internal states may be expressed in many different ways (J. K. Burgoon et al., 2009). Nevertheless, the nonverbal behavior expressed by people who try to hide their intentions in stressful situations might offer informative cues for security personnel. Until now, there is rather little empirical evidence for such nonverbal behavioral cues. Lie detection studies show rather ambiguous results and do not support typical nonverbal behavior cues to lying (Vrij, 2008). Liars as well as offenders planning a criminal act both try to deceive. However, the two kinds of deception differ for at least two reasons. Firstly, the typical build-up phase of a criminal act is missing in an interrogation setting: as soon as the interrogation starts, the liar already takes action (Frey et al., submitted). During the build-up phase of a criminal act, on the other hand, the actual intended action, for example the theft, has not yet taken place. Secondly, an interrogation includes an interaction with at least one interviewer who has to be persuaded. A build-up phase of a criminal act mostly includes no direct interaction with security personnel, but the suspects need to hide their intentions by becoming one with the crowd.

There is only very little research about possible nonverbal behavioral cues to criminal intentions during the build-up phase of a criminal act so far. Troscianko et al. (2004), for example, mentioned nonverbal behavioral cues that seem to be important for the correct prediction of aggressive behavior, such as *distinctive gaits*, *hand gestures*, *looking away from*

walking direction, and *looking around repeatedly*. Grant & Williams (2011) replicated and extended the findings by Troscianko et al. (2004) with eye tracking data. Their results indicate the importance of looking at the *face/head of individuals who were not engaged in reciprocal social interactions*. If individuals were engaged in social interactions, looking at their *bodies* was most successful. Heubrock et al. (2009b) and Heubrock (2011) conducted field studies in order to find out more about typical behaviors that reveal criminal intentions before assassinations. Heubrock et al. (2009b) asked participants to either try to "attack" and "murder" a person of interest with a realistic gun, or to give back a wallet to this person (control situation). The offenders had to become one with bystanders, act as inconspicuously as possible in order to reach their victim, and successfully solve their task. The authors conducted two experiments with this design. One of them took place in a small closed area, such as a classroom, whereas the other one took place in a foyer of a university, which was representative of a larger room with more bystanders. In the classroom situation, offenders showed a more *rigid behavior* in facial expressions, gestures, and the body than non-offenders. In the foyer situation, offenders displayed a higher frequency of *object-adaptors* (use of objects like cell phone without instrumental goal) and *self-adaptors* (movements with hands on own body to calm oneself), as well as behavior that is typical for *search and approach*. In addition, Heubrock (2011) conducted a field study concerning nonverbal behavior of possible suitcase bombers with Islamic background in a public train station. He asked three participants to place a suitcase bomb within 10 minutes at a place where the largest amount of destruction would be reached. Three travelers, who had agreed to be filmed while trying to catch their train, served as controls. The author again found a *higher frequency of object and self-adaptors*, as well as *more complex moving patterns* in offenders.

All these studies show promising results. However, additionally to the rather small sample size in the studies of Heubrock et al. (2009b) as well as Heubrock, (2011), all of the above mentioned studies suffer from the same shortcoming, namely, that different people

acted in the offender vs. non-offender situations. Thus, there is no control for inter-individual differences in behavior. In the studies mentioned above, the situations were kept similar, but the actors were different. In the worst case, people who always behave nervously might have been given the offender tasks, and people who are typically very calm might have been given the control task. This would result at least in an overestimation of the nonverbal cues. Therefore, it is very important to compare the same people in offender and non-offender situations.

Eachus et al. (2013) conducted a field study on stress symptoms and behavioral cues of hostile reconnaissance in public crowded spaces controlling for possible inter-individual differences. They examined four factors which could be of importance for behavior detection during the build-up phase done by humans as well as machines; a) cortisol and stress pheromones as biological factor, b) heart-rate, respiration, and body temperature as physiological factor, c) subjective measures of stress by the State Trait Anxiety Inventory (STAI, Marteau & Bekker, 1992) as psychological factor, and d) possible behavioral cues found through analyses of CCTV recordings of two of the 20 offenders by 15 other participants as behavioral factor. There were three conditions: (1) twenty mock offenders were sitting quietly in a room for 15 minutes (baseline), (2) the offenders were asked to traverse the shopping mall on a given route and to take pictures at six given locations (low intent condition), (3) the offenders were asked to take the same route and pictures again, but they had to be 10% faster than before and had to make sure they stayed undetected by plain-clothes security staff (high intent condition). The analyses of most measurements (Cortisol, Sterol, heart rate, respiration, and STAI) resulted in an increase in the high intent condition, which could not be fully explained by the higher physical activation in the high intent condition. Additionally, 15 participants analyzed the low and high intent videos of the person with the lowest STAI score and the person with the highest STAI score but found no differences in behavior. These results indicate on the one hand that laypeople indeed feel

different and stressed in mock crime situations, but, on the other hand, they contradict the findings mentioned before where some possible behavioral indicators for criminal intentions have been found (Troscianko et al., 2004; Heubrock et al., 2009b; Grant & Williams, 2011; Heubrock, 2011). Thus, it is not yet clear whether diagnostic nonverbal behavioral cues to criminal intentions can be found. The found ones may actually be behavioral artifacts due to the comparison of person A in situation A to person B in situation B. This would explain why Eachus et al. (2013) found no differences because they controlled for inter-individual differences. However, a shortcoming of this study was the rather small amount of recordings that was analyzed.

In general, the present study aims at clarifying these ambiguous results focusing on deceptive offenders such as mock thieves and mock terrorists. In order to control for inter-individual differences, we created our own recordings containing a mock crime and a control situation for every offender (team). The control situation consisted of a directed search for a (fictive) person (see method section for more details). In order to be able to compare the behavior of the same person in a mock crime and a control situation, we compared the recordings to these two conditions.

In Experiment 1 we took a closer look at the behavior of offenders regarding *moving pattern* in public space and *communication behavior*. In addition, we validated our own video recordings with regard to the two behaviors of interest by comparing the results of our own recordings with CCTV recordings containing real crimes (thefts).

We expected for all conditions (search, mock crime, real crime) a deviation in the *strength* of the behavior compared to the baseline, for *moving patterns* (Hypothesis 1a) as well as for *communication behavior* (Hypothesis 1b), since in all three conditions the persons of interest pursued an objective and tried to be successful. As criminals have to try to act as inconspicuously as possible in order to be successful, we expected this deviation being

smaller in mock as well as real crimes compared to the search condition for *moving patterns* (Hypothesis 2a) as well as for *communication behavior* (Hypothesis 2b), since trying to be as inconspicuously as possible can be defined as "trying to become one with the surrounding bystanders (= baseline)".

Besides the strength of behavior we were also interested in the *kind of behaviors* that are shown in the three conditions. We expected to find a difference in the kind of behavior in all three conditions for *moving pattern*s (Hypothesis 3a) as well as for *communication behavior* (Hypothesis 3b). For the same reasons as with the strength of behavior we expected this difference to be smaller for mock and real crimes compared to the search condition for *moving patterns* (Hypothesis 4a) as well as for *communication behavior* (Hypothesis 4b).

In Experiment 2 we took a closer look at *self-* and *object-adaptors*. We defined *self-adaptors* according to Sporer and Schwandt (2007) as "duration or frequency of movements in which one hand is in contact with the other hand or other parts of the body or face, such as by rubbing or scratching" (p. 11). *Object-adaptors* were defined according to Ekman and Friesen (1972) as well as Heubrock (2011) as the use of an object rather in order to calm oneself than for its instrumental goal. We examined these two adaptors in detail based on the findings by Heubrock (2011) and Heubrock et al. (2009b), but also because of our personal conversations with criminal investigators. These personal conversations revealed hypotheses about positive correlations between *self-adaptors* and *object-adaptors* and criminal intentions. Similar assumptions of laypeople and professionals are found in lie detection research for *self-adaptors* (Mann et al., 2004; The Global Deception Research Team, 2006; Sporer & Schwandt, 2007), although, in reality no correlations were found between lied statements and *self-adaptors* indicating that *self-adaptors* might not be a reliable cue (DePaulo et al., 2003). Research on lie detection is usually conducted in interrogation settings (Vrij, 2008). In addition to the above mentioned differences this setting differs from the setting during the

build-up phase of a criminal act as truth-tellers might also use strategies in order to persuade the interviewer not to mistake them for liars (DePaulo et al., 2003; Strömwall et al., 2006). Thus, the fact that no correlation between *self-adaptors* and telling a lie has been found might be due to the strategic behavior of truth-tellers and liars on the one hand, and/or the methodological problem of having different participants in truth vs. lie situations. During the build-up phase of a deceptive criminal act, such as a theft or placing a bomb, these correlations might be different from situations used in lie detection research for two reasons. First, we compare the same people once having a hidden criminal intention and once having a goal-driven but noncriminal intention. Thus, for every mock offender, we can differentiate nonverbal behavioral cues due to criminal intentions from nonverbal behavioral cues in situations with a goal-driven but noncriminal intention. Second, our search (control) condition differs from the truth condition in interrogation settings insofar that there is no need for persuading for example an interviewer of something. Thus, in the search (control) condition there is no need for any alteration of people's behavior. Based on the above mentioned hypotheses found in personal conversations, we, therefore, expect stronger presence of *self-* (Hypothesis 5a) and *object-adaptors* (Hypothesis 5b) in the mock crime condition than in the control search condition[1].

3.2. General method

3.2.1 Used scenarios

Two different scenarios, namely a *search* condition (control situation), and a *mock crime* condition, were recorded on video. These recordings served as basic material for both experiments. Laypeople were chosen as "actors" in these situations based on the results by Eachus et al. (2013) mentioned in the introduction. Since different people show different

[1] The rather low quality and the distance view of CCTV recordings containing real crimes did not allow for using them in Experiment 2. Therefore, no hypothesis could be tested in this context although we would as in Experiment 1 expect no difference between mock and real crimes.

nonverbal behavior (Dente et al., 2006), we had each actor or pair of actors acting in both conditions. It is important to note that these actors were not instructed to play act, but that they believed to be part of training for criminal investigators (see cover story) in both conditions.

The *search (control) condition* consisted of the task to search for another person. The *mock crime condition* consisted of the task to either commit a *theft*, or to place a *mock bomb* in such a way that a target victim or as many people as possible would be killed. In both sorts of crimes, there is a build-up phase during which the offenders try to hide their criminal intentions. The conditions were staged in three public areas, namely a large international airport, a large train station, or a public interchange place in order to account for possible differences in behavior due to the location where the mock crime is committed.

3.2.2 Actors and cover story

Thirteen security control officers at a large international airport acted as mock offenders. Ten of them operated in teams of two, three of them acted alone. They were told they would participate as mock offenders in a training for criminal investigators, where the criminal investigators have to observe a crowd to find possible offenders without being identified as criminal investigators themselves. Before the mock offenders were instructed about their task, the search condition was recorded in which they were told to search another (non-existing) instructor in order to get more information for the (later mock crime) scene area. In order to receive enough material this task always lasted at least five minutes and maximally ten minutes. In this *search (control) condition*, the mock offenders did neither try to hide any criminal intention nor act as inconspicuously as possible. We decided to use a search instead of a "doing nothing" situation in order to make sure that there is a clear goal in every condition and in order to later be able to distinguish the behavior due to hiding of criminal intentions from mere active search behavior.

For the *mock crime condition*, we instructed our mock offenders a second time and told them that they should try to deceive the criminal investigators that were to be trained. Thus, the mock offenders had to commit the crime in such a way that the investigators would not be able to identify them as the offenders. They were given ten minutes to go into the situation, commit their crime, and leave the situation again. Our actors were not aware of the fact that no criminal investigators were there to be deceived. The "victims" in the mock crime condition consisted of laypeople that were told that they would participate in a field study on attention in real world scenarios. They were instructed to read a story from a children's book in a crowded and noisy public place and that they later would have to answer questions about the book's content. Thus, they were distracted by a cognitive task as well as by the crowded place, so that their ecological validity as mock victims was given.

All actors were informed beforehand that they will be recorded on video, and written consent to the study and the recordings was obtained from them. They were also made aware of the possibility to withdraw their written consent at any time. After completion of their tasks, all actors received a thorough debriefing about the main objective of the study.

3.2.3 Video recordings

All situations were recorded by two overview cameras (further called *space* condition) and three to five cameras which zoomed in on the offenders (further called *zoom* condition). In order to obtain zoom recordings in which the mock offenders are always visible, a naïve person cut out the parts where the camera lost focus, where the mock offender walked out of the image, or where public transportation drove in between the mock offender and the camera. The duration of these recordings varied due to differences in the duration of the build-up phase, which was dependent on the mock offenders' way of committing the mock crime (*space* condition: M = 196.13 s; SD = 94.78 s). Some of the recordings of the zoom condition were shortened due to the above-mentioned cutting (*zoom* condition: M = 175.15 s; SD =

83.09 s). Each mock crime recording ended shortly before the mock crime was committed, thus only showing the build-up phase. The recordings of the search (control) conditions were cut to the same length as the according mock crime recordings. We used eight packages of recordings; detailed information about the mock crimes, number of offenders, duration and locations is shown in *Table 1*.

Table 1
Own video recordings: type of mock crime, number of offenders, duration, and location. Note that duration and location and number of offenders were the same for the search (control) condition.

	Mock crime	Offenders	Duration in s for space and zoom (both offenders)	Location
Package 1	Bomb	1	Space: 143 s Zoom: 87 s	Airport
Package 2	Theft	2	Space: 209 s Zoom: 173 s / 193 s	Airport
Package 3	Theft	2	Space: 93 s Zoom: 134 s / 94 s	Train station
Package 4	Bomb	1	Space: 86 s Zoom: 95 s	Train station
Package 5	Theft	2	Space: 180 s Zoom: 134 s / 103 s	Train station
Package 6	Bomb	1	Space: 243 s Zoom: 162 s	Train station
Package 7	Theft	2	Space: 288 s Zoom: 188 s / 262 s	Public interchange station
Package 8	Theft	2	Space: 363 s Zoom: 315 s / 336 s	Public interchange station

As a complement to the recordings with actors, we used real CCTV recordings of baggage thefts at a large international airport. These were cut shortly before the theft was committed so that they again only showed the build-up phase of the crime. The length of each recording was defined by the offenders' way to commit the crime (M = 87.20 s; SD = 38.22 s). For a detailed description of these CCTV recordings, see Frey et al. (submitted).

3.3 Experiment 1 – Moving patterns and communication behavior

In Experiment 1, we were interested in *moving patterns* in public space as well as nonverbal *communication behavior* in general. Based on personal conversations with criminal investigators we were interested in *abrupt changes in direction* or *speed*, and *position changes* as behavioral cues relating to *moving patterns*. We furthermore analyzed *hand signs*, *hand movements, use of cell phone, lip movements,* and *gaze movements* as nonverbal behavioral cues for *communication behavior*.

3.3.1 Participants

Eighteen undergraduate students of the University of Zurich participated. Seventeen participants were female and one was male, with an average age of 24.56 years (*SD* = 6.60). They were either paid according to the usual conditions at the University of Zurich (i.e., CHF 15 per hour) or received course credits. All participants were naïve to the actual background of the experiment, as experience and expectations can influence the perception of nonverbal behavior (Levine, Asada, & Park, 2006; Fuhrer, 2013). At the end of the experiment, participants were debriefed thoroughly about the reasons of the study. They were also informed of the possibility to withdraw their consent at any time.

3.3.2 Stimuli

In this experiment, we only used the recordings of the *space* condition. Thus, there were ten authentic CCTV recordings of real baggage thefts. Additionally, we used the eight space condition recordings of the mock crimes and the according eight searches. This resulted in a total of 26 stimuli which were divided into two sessions with 13 stimuli each. If a participant saw the mock crime recording of Person A in session 1, he/she saw the search of Person A in session 2 (Version A), and vice versa (Version B). Within one session, all recordings were presented in random order and each session contained five of the ten CCTV

recordings. Half of the participants saw version A, the other half saw version B. Participants were randomly assigned to these versions.

3.3.3 Procedure

Participants were asked to participate in this experiment on two different days. Each session lasted one hour. The experiment combined computer-based and paper-pencil methods. Instructions and recordings were presented on a computer using the java-based open-source programming framework *Tatool* (von Bastian, Locher, & Ruflin, 2013). The participants responded on paper to give them the possibility to easily re-check the list of relevant behaviors of interest while watching the recordings.

At the beginning, the participants were informed that they would participate in an experiment about nonverbal behavior in public places. They were told that they were going to watch video recordings, for which they would assess nonverbal behavior of pre-defined persons of interest. The detailed definition of the relevant nonverbal behaviors of interest that had to be judged was followed by an instruction check. In the instruction check, participants were shown two different recordings containing simple examples of the behavior of interest. After each recording, they had to indicate which of the two behavior groups (either *moving patterns* in public space or *communication behavior*) they observed. Performance was reviewed by the investigator: if the answers were not correct, more instruction would have followed, which never happened. After this instruction check, the concept of the *baseline* was explained as "How strongly do the other people in the recording show this behavior? How strongly would you expect this behavior to be shown in this situation?" Then, the response sheet was explained in detail. Participants had to indicate separately for each behavior group (*moving patterns* and *communication behavior*) on a visual analogue scale: a) how strongly the behavior was shown by the other people (baseline), b) how strongly the behavior was shown by the people of interest, c) how strongly the *kind* of behavior shown by the people of

interest differed from the kind of behavior shown by the baseline. In addition, participants had to indicate which of the following behaviors they had observed for *moving patterns* (*abrupt changes of direction, abrupt changes in speed, changes of position*, own observations) and for *communication behavior* (*hand signs, head movements, use of cell phone or phone booth, lip movements, gaze movements*, own observations). We instructed them that the explicitly mentioned kinds of behavior were not exclusive for the behavior of interest groups. Thus, they were instructed to note what other behaviors they observed and found notable in the category *own observation*. After the instruction was finished, an example recording was shown to give participants the possibility to get used to the procedure.

The first frame of each recording was presented as a still, in which the people of interest were marked with red circles. The recording started automatically after eight seconds. A slide was presented at the end of each recording, which asked the participants to give their responses. At the end of the first session, the participants were asked to participate in the second session. At the end of the second session, the participants were thoroughly debriefed about the background and the research questions of the experiment, the mock crimes, and the searches.

3.3.4 Results

We standardized the answers of all visual analogue scales into values from 0 (behavior was not shown at all) to 100 (behavior was maximally shown), in order to express deviations in terms of percentage. In order to test hypothesis 1a and 1b, we calculated the deviation from the baseline separately for each behavior of interest (*moving patterns* and *communication behavior*), i.e., we subtracted the standardized value that was indicated for the baseline from the standardized value for the person(s) of interest for each participant and each recording. As we were not primarily interested in the direction of the deviation from the baseline, i.e., whether the person(s) of interest showed the behavior of interest more or less strongly than

the people who formed the baseline, but rather in the question whether the strength of the behavior of interest differs at all we used the absolute values for statistical analyses. Based on these absolute values, we calculated one-sample T-tests for each behavior of interest in order to analyze whether the deviation from the baseline is significant. In all conditions, all judged behaviors deviate significantly from the baseline, which is zero in our case (*moving patterns* and *communication behavior*: $p < .001$, for all conditions). This indicates that in all conditions, the strength of the behavior of interest was consistently perceived to differ for the person(s) of interest compared to the baseline. Thus, hypotheses 1a and 1b are confirmed.

To test hypotheses 2a and 2b, we calculated two separate analyses of variance (ANOVAs) for *moving patterns* and *communication behavior* with condition (search, mock crime, and real crime) as within-subjects variable. The respective data is displayed in *Figure 1*.

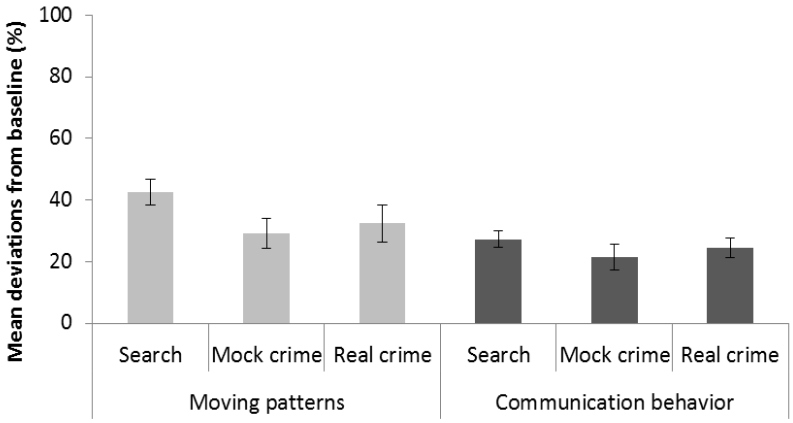

Figure 1. Mean absolute differences in % between the behaviors observed in the person(s) of interest and the baseline (zero). Error bars represent confidence intervals (95%) for within-subject comparisons calculated according to Cousineau (2005) and Morey (2008).

Pairwise comparisons are always adjusted for multiple comparisons according to Bonferroni. The ANOVA for *moving patterns* showed a significant effect of condition, $F(2, 34) = 7.234$, $p = .002$, $\eta^2 = .298$, indicating a difference in the absolute value of deviation from the baseline for at least one of the conditions. The pairwise comparisons confirmed significant differences between mock crime and search ($p = .001$), and real crime and search

(p = .046). *Figure 1* shows that the deviance from the baseline is smaller in the two crime conditions compared to the search condition, confirming hypothesis 2a. There was no significant difference between mock crimes and real crimes (p = 1.000), thus underpinning the high external validity of the mock crime recordings. The ANOVA for *communication behavior* revealed no significant effect of condition, $F(2, 34) = 2.734$, $p = .079$, $\eta^2 = .139$. Thus, Hypothesis 2b cannot be confirmed.

A descriptive analysis indicates how many percent of the difference values were negative, and how many percent of the difference values were positive for the behavior of interest as compared to the baseline. *Figure 2* reveals that most difference values were positive, i.e., the behavior of interest was more strongly expressed by the person(s) of interest as compared to the baseline. For *communication behavior*, the pattern appears to be less uniform, particularly in the condition "real crime".

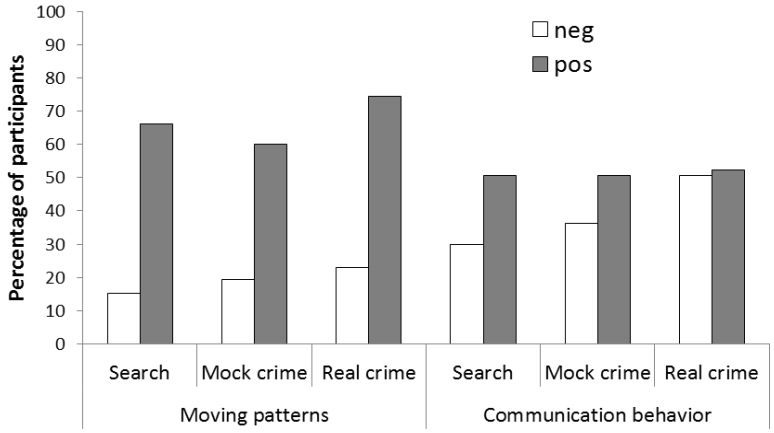

Figure 2. Percentage of participants who reported positive vs. negative deviation from the baseline for each behavior of interest (*moving patterns* and *communication behavior*) and condition (search, mock crime, and real crime).

Figure 3 shows the averaged values expressed in terms of percentage for the answers to the question "how strongly differed the kind of behavior observed in the person(s) of interest from the kind of behavior observed in the baseline?" from zero (not at all) to 100 (very strongly) for each behavior of interest (*moving patterns* and *communication behavior*) and condition (search, mock crime, and real crime). One-sample T-tests with these values

showed significant differences from zero for both behaviors of interest (*moving patterns* and *communication behavior*) in all conditions ($p < .001$), thus confirming hypotheses 3a and 3b.

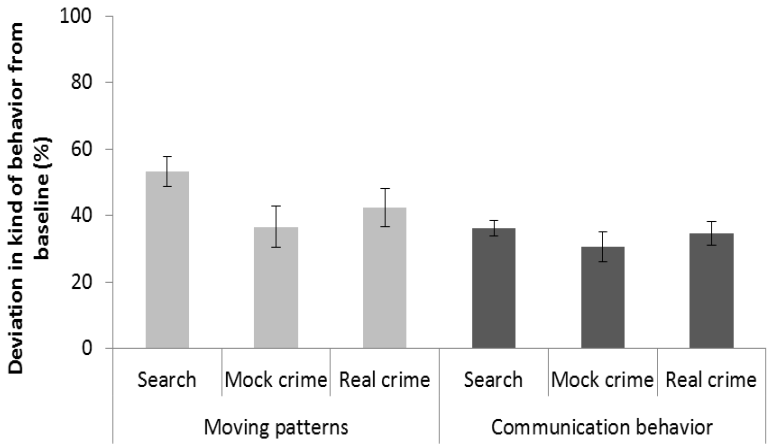

Figure 3. Mean values (%) indicating how strongly the kind of behavior observed in the person(s) of interest differed from the kind of behavior observed in the baseline for each behavior of interest (moving pattern and communication) and condition (search, mock crime, and real crime). Error bars represent confidence intervals (95%) for within-subject comparisons calculated according to Cousineau (2005) and Morey (2008).

The ANOVA for *moving patterns* with the data displayed in *Figure 3* showed a significant effect of condition, $F(2, 34) = 9.067$, $p = .001$, $\eta^2 = .348$, indicating that the answers regarding the kinds of behavior differed between at least two conditions. Pairwise comparisons showed significant differences in *moving patterns* being between mock crime and search ($p = .001$), and real crime and search ($p = .018$). There was again no significant difference between mock crime and real crime ($p = .663$). Based on these results and the data displayed in *Figure 3*, hypothesis 4a can be confirmed. In contrast, the ANOVA for *communication behavior* showed no significant effect of condition, $F(2, 34) = 2.380$, $p = .108$, $\eta^2 = .123$. Thus, hypothesis 4b is not confirmed.

For explorative purposes, we looked more closely at the different kinds of behaviors that were indicated by the participants. As dependent variable, we calculated the percentage of how often per condition the box of each behavior was checked. For *communication behavior* we excluded the behaviors *lip* and *gaze movements* in the condition real crime, because they were actually not visible due to the low quality of CCTV recordings. All behaviors of interest

that could have been mentioned by the participants are displayed in *Table 2* and the resulting data is displayed in *Figure 4a* for *moving patterns* and *Figure 4b* for *communication behavior*. It has to be kept in mind that these results only cover responses to the behaviors that we had indicated on the questionnaire.

Table 2
Examples for the kind of behavior included in the concepts used as behaviors of interest for Experiment 1. Note that these behaviors were included in the response sheet with an added category "own observations", see Table 3. These categories were not exclusive.

Behaviors of interest	
Moving patterns	Communication behavior
Abrupt/many direction changes	Hand signs
Abrupt/many speed changes	Head movements
Many position changes	Use of phone booths / cellphones
	Lip/mouth movements
	Gaze movements

Participants could note further behaviors they found notable in the category *own observations*. For the sake of completeness, these notes are displayed in *Table 3.*

Table 3
Summarized notes of participants in category "own observations" for moving patterns and communication behavior split per condition (search, mock crime, and real crime).

Behaviors of interest					
Moving patterns			Communication behavior		
Search	Mock crime	Real crime	Search	Mock crime	Real crime
Remove jacket	Knows where to go	From moving systematically to moving in a confused way	Looks down	Approach others	Use of Check-in station
Hands in face/hair	Sitting	Puts on hood	Empty look	Communicates with strangers	Takes baggage trolley without real reason
Reading schedule	Hands in pockets	Waiting	Takes pictures	Smells flowers	
Hands on hips	Leaning arm on railing	Watches baggage of strangers	Don't know where to go	Large distance between team members	
Lean on railing	Smoke cigarettes	Smoke cigarettes	Interaction	Talked a lot to each other	
To point to sth	Small distance towards strangers	Turn around on place	Searching		
Few movements	Throws sth into the garbage	Systematical position changes	Seems as if they don't know each other		
Waiting	Strolling	Searches hectically	Folding their arms		
Pause	Changes position systematically	Moving fast from left to right and back	Distance communication		
Confused	Searching	Takes hands behind back	Touches other's shoulder		
Bored	Turn around on place	Searches	Touches other people		
Squeeze between suitcases	Studying plates	Squeeze between suitcases	Looks around all the time		
Distance between team member changes	Confused				
Tapping the floor by foot	Moves fast				
	Separate and come together				

We calculated an ANOVA for each behavior that had been indicated explicitly on the questionnaire in both behaviors of interest groups. These ANOVAs featured condition (search, mock crime, real crime) as independent variable. Afterwards, pairwise comparisons (Bonferroni) were calculated.

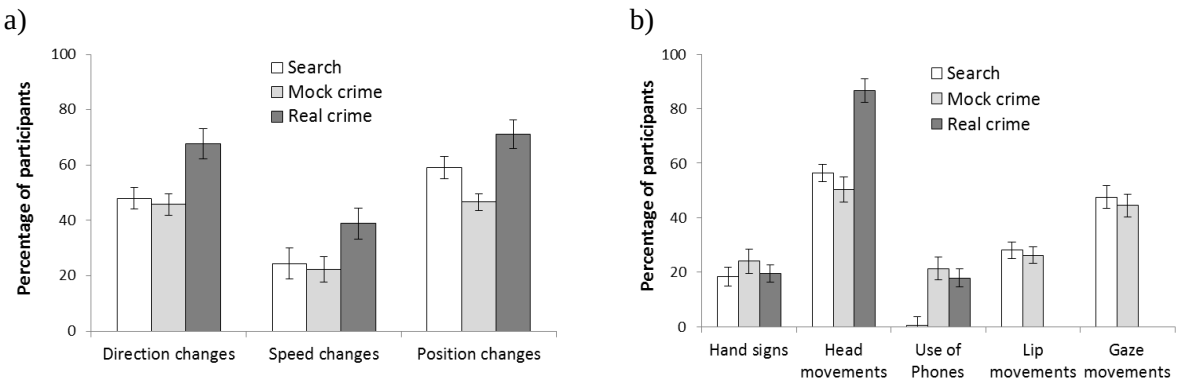

Figure 4. Percentage of participants who observed the according kind of *moving pattern* (a) and *communication behavior* (b) per condition. Error bars represent confidence intervals (95%) for within-subject comparisons calculated according to Cousineau (2005) and Morey (2008).

The ANOVA for *direction changes* showed a significant effect of condition, $F(2, 34) = 27.872$, $p < .001$, $\eta^2 = .621$. Pairwise comparisons showed a significant difference between real crime and search as well as mock crime (both $p < .001$), and no significant difference between mock crime and search ($p = 1.000$). The ANOVA for *speed changes* showed a significant effect of condition as well, $F(2, 34) = 11.198$, $p < .001$, $\eta^2 = .397$. Pairwise comparisons showed a significant difference between real crime and mock crime ($p = .001$), and real crime and search ($p = .010$). There was no significant difference between mock crime and search ($p = 1.000$). The ANOVA for *position changes* showed a significant effect of condition, $F(1.486, 25.255) = 32.435$, $p < .001$, $\eta^2 = .656$ (Greenhouse-Geisser). Pairwise comparisons showed a significant difference between all conditions (mock crime vs. search: $p < .001$; mock crime vs. real crime: $p < .001$; real crime vs. search: $p = .014$). The ANOVA for *hand signs* showed no significant effect of condition, $F(2, 34) = 2.564$, $p = .092$, $\eta^2 = .131$. The ANOVA for *head movements* showed a significant effect of condition, $F(2, 34) = 88.880$, $p < .001$, $\eta^2 = .839$. Pairwise comparisons displayed a significant difference between real crime and search as well as mock crime (both: $p < .001$), but no significant difference between mock crime and search ($p = .118$). The ANOVA for the *use of cellphone / phone booths* showed a significant effect of condition, $F(2, 34) = 37.630$, $p < .001$, $\eta^2 = .689$. Pairwise comparisons showed a significant difference between search and mock crime as well

as real crime (both: $p < .001$), but no significant difference between mock crime and real crime ($p = .685$). Pairwise comparisons for *lip movements* ($t(17) = -.809$, $p = .430$) and *gaze movements* ($t(17) = -1.022$, $p = .321$) showed no significant differences between mock crime and search.

3.4 Experiment 2 – Object-adaptors and self-adaptors

In this experiment, we focused on *self-adaptors* (Sporer & Schwandt, 2007; Heubrock, 2011) and *object-adaptors* (Ekman & Friesen, 1972; Heubrock, 2011). We examined whether *self-adaptors* and *object-adaptors* are expressed differently by the same person when performing a goal-driven search without criminal intention vs. pursuing a criminal goal, the latter requiring the participants to try to act as inconspicuously as possible.

3.4.1 Participants

Sixteen participants with an average age of 24 years ($SD = 4.425$) took part in this study. Ten of them were undergraduate students of psychology, one had already finished the master's degree in psychology. Five participants were no students and worked outside the university. Twelve participants were female. They were all naïve to the actual goal of the study for the same reason as in Experiment 1. All participants were either paid according the usual conditions of University of Zurich (15 CHF/h) or received course credit after having participated in both sessions. They signed an informed consent and were aware of the possibility to withdraw from the study at any time.

3.4.2 Stimuli

The same eight recording packages as in Experiment 1 served as stimuli. In contrast to Experiment 1, only the *zoom* condition was used. Therefore, every situation of every offender was included, which resulted in a total of 26 recordings (5 offender teams = 10 searches, 10

mock crimes; 3 single offenders = 3 searches, 3 mock crimes). These recordings were distributed into two sessions in a quasi-randomized way insofar that one of the two recordings of each offender was shown in session 1 and the other in session 2. Additionally, we made sure that approximately the same amount of mock crime and search recordings were displayed in every Block (Block A: 7 searches, 6 crimes; Block B: 6 searches, 7 crimes). The Block order was counterbalanced insofar that half of the participants saw Block A in session 1 and Block B in session 2; the other half saw it vice versa.

3.4.3 Procedure

All participants were asked to participate in the experiment on two different days. The procedure was the same as in Experiment 1. The detailed instruction about the task was again displayed on the computer, and the participants answered on paper. For the instruction check, we again showed them recordings with examples of *object-adaptors* and *self-adaptors*. After each recording, participants had to indicate which behavior of interest was shown. The experimenter observed their answers and explained the task and the concepts in more detail if needed. Contrary to the recordings in Experiment 1, bystanders were not visible, and neither did participants have a full and clear view of the area. Thus, participants could not establish a baseline. Therefore, participants had to give their answers only on one visual analogue scale per behavior of interest, namely: "how strongly was this behavior shown by the person of interest?". In addition to their answer on the visual analogue scale, participants could indicate which kind of explicitly asked *self-adaptors* or *object-adaptors* they observed (see *Table 4* for details) using check boxes. In addition, they could note additional nonverbal behavior that they found notable in the category *own observations* (see T*able 5* for details).

Table 4
Examples for the kind of behavior included in the concepts used as behaviors of interest for Experiment 2. Note that these behaviors were included in the response sheet with an added category "own observations", see Table 5. These categories were not exclusive.

Behaviors of interest	
Object-adaptor	Self-adaptor
Use of Cellphone without instrumental goal	Hand movements in the face
Actions performed on own things to calm oneself (e.g. dig in bag)	Hand movements in the hair
Actions performed on external objects to calm oneself	Hand movements on whole body
	Fidgeting
	Hand wringing
	Touch and scrub one's own clothing

Table 5
Summarized notes of participants in category "own observations" for self-adaptors and object-adaptors split by condition (search, mock crime).

Behaviors of interest			
Self-adaptors		Object-adaptors	
Search	Mock crime	Search	Mock crime
Hands often in pockets	Pressing lips together	Studies map	Actions on other person
Mouth twitches	Hands in pockets	Plays around with bag	Chewing gum
Clenching one's fists	Wringing fingers	Leans over railing	Observes other people
Stretches the neck	Scratching wrist	Tapping the floor by foot	Use of phone booth
Fidgeting on jacket	Playing with fingers	Plays around with ring	Checking watch
Scratches hands	Moving around	Holds jacket	Drinking a lot
Holds jacket	Scratching the head	Puts bag down next to himself	Fidgeting on bag
Hands on collar	Folding his arms		Cigarettes
Making gestures	Making gestures		Tries out phone booth
Bites on lips	Drawing on hand		Holds the railing
Folded arms moved			Studies schedule / map
Folds hands behind back			Reads the newspaper
Scratching			

3.4.4 Results

As in Experiment 1, we first standardized the answers given on the visual analogue scales into values from zero (not at all) to 100 (very strongly) in order to express deviations in terms of percentage. After standardization, we calculated a *difference value* for each person of interest between his/her mock crime and search (control) condition. Thus, we subtracted the value in the search (control) condition from the value in the mock crime condition for every

person of interest. Contrary to Experiment 1, we were interested in the direction of deviation, because we hypothesized that *object-adaptors* as well *as self-adaptors* would be more strongly shown in the mock crime condition as compared to the search (control) condition. *Figure 5* displays the average values over all persons of interest and all participants.

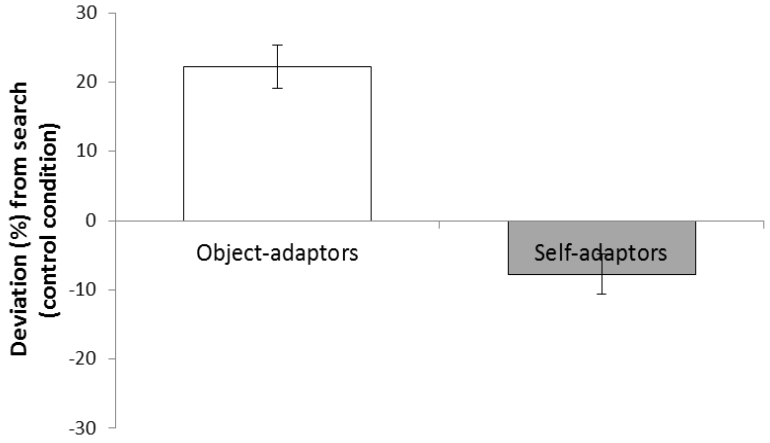

Figure 5. Deviation (%) in frequency of self-adaptors and object-adaptors from the control search situation (zero). Positive values indicate more expression in the mock crime situation, whereas negative a value indicate stronger expression in the control search situation. Error bars represent confidence intervals (95%) for within-subject comparisons for each behavior of interest calculated according to Cousineau (2005) and Morey (2008).

First, we calculated a one-sample T-test against zero for each behavior of interest. This test showed a significant difference for *object-adaptors* ($t(15) = 9.865$, $p < .001$) as well as *self-adaptors* ($t(15) = -3.681$, $p = .002$). The data in *Figure 5* reveal a positive deviation from zero for o*bject-adaptors* and a negative deviation from zero for self-adaptors, which partially hypothesis 5a: *object-adaptors* were indeed expressed more strongly in mock crime than in search (control) condition. However, hypothesis 5b could not be confirmed: *self-adaptors* were expressed less strongly in the mock crime condition than in the search (control) condition.

Second, we were interested in what *kind* of behavior the participants actually observed. Again, it has to be kept in mind that this question was independent of the first one about how strongly the behavior was shown, because the given kinds of behavior were not exclusive. For the sake of completeness, *Table 5* lists the kinds of behavior that participants

additionally noted in the category *own observations*. For the explicitly stated kinds of behavior (see *Table 4* for details*)*, we calculated for each condition the percentage of participants who mentioned that kind of behavior of interest. These data are displayed in *Figure 6a* for *object-adaptors* and *Figure 6b* for *self-adaptors*.

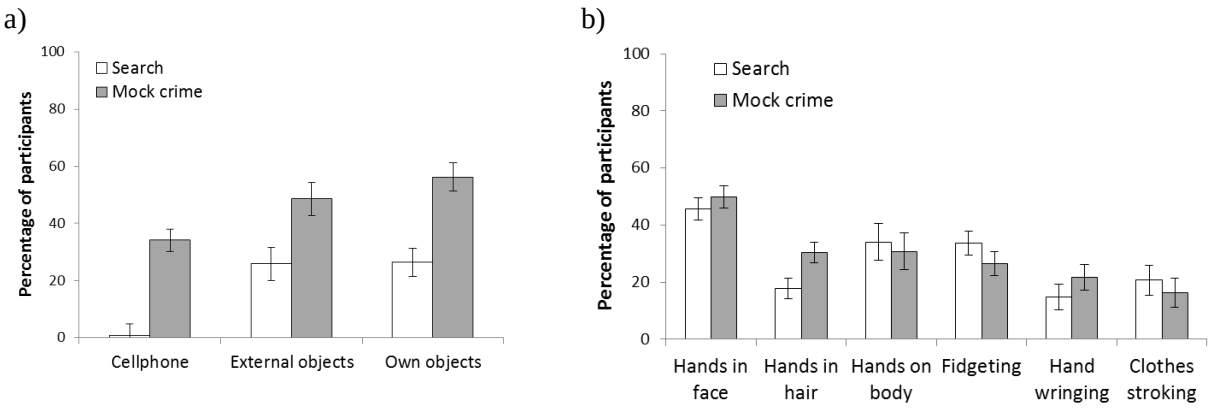

Figure 6. The percentage that the according kind of *object-adaptor* (a) or *self-adaptor* (b) was observed by participants in each condition. Error bars represent confidence intervals (95%) for within-subject comparisons, calculated according to Cousineau (2005) and Morey (2008).

We subsequently calculated paired T-Tests for each kind of behavior of interest. The detailed statistics are displayed in *Table 6*. For these multiple comparisons, we lowered the significance level to $p < .01$ according to Bonferroni. The calculations resulted in a significant difference between mock crime and search for all kinds of *object-adaptors* and for the self-adaptor *hands in hair*. The evaluated *object-adaptors* as well as the *self-adaptor hands in hair* were all judged to occur more strongly in the mock crime condition, all other comparisons were not significant.

Table 6
T-tests for the use of object-adaptors and self-adaptors in mock crime vs. search (control) conditions.

	Pairs mock crime vs. search	T	df	Significance (two-sided, Bonferroni)
Object-adaptors	Use of cellphone without instrumental goal	11.915	15	.000*
	Actions performed on own things to calm oneself	6.376	15	.000*
	Actions performed on external objects to calm oneself	7.188	15	.000*
Self-adaptors	Hand movements in the face	1.542	15	.144
	Hand movements in the hair	4.779	15	.000*
	Hand movements on whole body	-.731	15	.476
	Fidgeting	-2.390	15	.030
	Hand wringing	2.049	15	.058
	Touch and scrub one's own clothing	-1.165	15	.262

3.5 Discussion

The main objective of this study was to investigate whether nonverbal behavioral cues can be identified that are related to hidden criminal intentions during the build-up phase. These nonverbal behavioral cues can consist of *moving patterns*, *communication behavior*, *self-adaptors*, and *object-adaptors*. Our study is the first one to use a combination of recordings of real crimes and own recordings of mock crimes and a search (control) condition for investigating the above research question. The use of this material permitted us to control for inter-individual differences as well as to demonstrate the external validity of our own recordings. Searching for a fictive person as control condition made it possible to later exclude nonverbal behavioral cues to mere search behavior (i.e., criminals need to search the perfect victim or place) from nonverbal behavioral cues to hidden criminal intentions. The results of this study indicate that offenders differ significantly from non-offenders in their nonverbal behavior.

In our first experiment, we found a significant deviance from the baseline for both *moving patterns* and *communication behavior* in mock crimes, real crimes, as well as in the search (control) condition. In other words, independent of the intention the persons of interest had, their nonverbal behavior deviated from the behavior of bystanders.

Heubrock (2011) found in his field study a *more complex moving pattern* of offenders than non-offenders (regular travelers). He describes the moving patterns for regular travelers as more straightforward and less complex, which would be in line with our interpretation. Our results complement the results by Heubrock (2011) by showing that *moving patterns* for people actively searching someone deviate even more strongly from the baseline than *moving patterns* of offenders. This finding is important as it shows that the premise "the stronger the deviation in moving pattern the more suspicious" is clearly wrong. Thus, a strong deviation in *moving patterns* can also be expressed by people without any criminal intentions. It may even be hypothesized that the (mock) offenders successfully tried to adapt their *moving patterns* to the baseline in order to convey the impression of being regular travelers. They may have successfully tried masking their real intentions by a behavior that appears as inconspicuous as possible. Thus, they may have tried to become one with the crowd of bystanders. This is in line with the already known fact that people try to control their behavior while trying to deceit (Ekman & Friesen, 1969b; Ekman, 1981; J. K. Burgoon & Buller, 1994; Vrij, 2008). Nevertheless, their criminal intentions "leaked" and expressed themselves in an observable deviation in moving patterns from the baseline. We did not find any differences between the three conditions for communication behavior, i.e., their deviations from the baseline are comparable in size. This rather unexpected result might stem from the control condition, which consisted of a goal-driven active search. A team of actors actively searching somebody might need the same amount of communication as offenders searching for the perfect victim.

We did not find the same pattern in the qualitative data on the observed *moving patterns*. Because we found no difference between mock and real crime in *moving patterns* in the comparison of person(s) of interest and baseline, we would have expected to find no difference between real and mock crimes concerning the explicit kinds of behaviors too. However, there was always a significant difference between real and mock crimes as well as searches for all three explicit kinds of behaviors. This result might be an artifact of the situational area since, contrary to the difference value for the *moving patterns*, these data are independent from the actual baseline. All CCTV recordings, and, therefore, all real crimes were recorded at a large international airport, but only two of our own recording packages were recorded at the same airport. Additionally, the crimes in the CCTV recordings were thefts exclusively, in our own recordings the crimes varied between thefts and terrorist acts. It might be the case that some explicit behaviors are more often expressed at an airport than for example at a public interchange station due to the architectural circumstances of the place (e.g., Hoogendoorn & Bovy, 2004). Thus, these explicit behaviors would automatically be observed more in CCTV recordings than in others, because the underlying baseline would already be higher in this situational area. For example, at an airport, bystanders might express more *abrupt direction* and *speed changes* as well as *position changes* than at a train station because of the architectural circumstances or the fact that people are less familiar with the airport. Thus, to become one with the crowd, offenders might as well try to express these behaviors more often at an airport than for example at a train station.

Alternatively, the different kinds of crimes could also lead to a different kind of explicit kind of moving pattern. For example, committing a theft might need more *abrupt direction changes* and *speed changes* than placing a bomb since, in the first case, the offenders might need to approach the victim more closely in order to steal. Thus, the fact that these explicit moving patterns were mentioned more often in the real crime condition could also be due to either architectural circumstances or the nature of the crime, or a combination

of both. However, as soon as the whole cluster of these nonverbal behaviors is taken into account, the amount of deviation of the *moving patterns* of real offenders and mock offenders does not differ anymore: The distribution of negative and positive deviations from the baseline looks similar for mock crimes and real crimes (*Figure 2*). This underlines the high importance of using the concept of a baseline and clusters of behaviors when judging other people's behavior.

In Experiment 2, we took a closer look at adaptors while controlling for inter-individual differences. *Self-adaptors* (e.g., Sporer & Schwandt, 2007) and *object-adaptors* (Heubrock et al., 2009b) are often believed to be reliable cues to deception when lying is concerned. Both were found with a higher frequency in offenders than non-offenders in the field studies by (Heubrock et al., 2009b) and Heubrock (2011). However, these results might be artefacts of the comparison of different people in the mock crime and control situations. With the use of our search (control) condition, we not only excluded possible inter-individual differences but also nonverbal behavioral cues due to the search part of the criminal act. We found in our study significantly more *object-adaptors* and less *self-adaptors* in the mock crime condition compared to the search (control) condition.

Contrary to our expectations, *self-adaptors* are expressed less frequently in the mock crime situations compared to the search (control) condition. Since lie detection research (Mann et al., 2004; The Global Deception Research Team, 2006; Sporer & Schwandt, 2007; Vrij, 2008) and our personal conversations with experienced police officers showed that professionals as well as laypeople believe that higher frequented self-adaptors are a sign of deception, we expected the actors to express more *self-adaptors* as offenders than as non-offenders.

A possible reason for this effect could lie in one of the three factors of the model by Zuckerman et al. (1981), namely attempted behavioral control. The underlying idea is that

liars try to control their nonverbal behavior to appear truthful or at least to not express any cues to their intention of deceiving. The results of our first experiment concerning *moving patterns* showed less deviation from the baseline in the crime conditions. Thus, it can be assumed that offenders tried to become one with the crowd and tried to act inconspicuously. Consequently, they tried to control their behavior. Since the offenders in Experiment 2 were exactly the same people in the same situation, it can be assumed to be true also for the expressed behaviors in this experiment. As the correlation between *self-adaptors* and nervousness as well as deception is a common belief in laypeople (e.g., Sporer & Schwandt, 2007), we argue that our offenders were aware of the possibility to express *self-adaptors,* and that they believed that such behavioral cues would let them appear suspiciously. Thus, the emotional factor of Zuckerman et al.'s (1981) three factor model interacted with the attempted behavioral control factor. The result was that our mock offenders presumably tried hard to control their behavior that expresses their higher arousal. In other words, they tried to suppress *self-adaptors* (Sporer & Schwandt, 2007). However, if that is the case, why did they just suppress *self-adaptors* but not *object-adaptors*? *Object-adaptors* are not well known up to the present. Even though they have already been mentioned by Ekman and Friesen (1972), only the Heubrock group has studied them in more detail as far as we know. For that reason, offenders might not be aware of them being a cue. Consequently, they do not try to suppress them. Moreover, *object-adaptors* might let offenders feel that they become one with the crowd: at least, they seem to be engaging in something. Besides, *object-adaptors* might still counteract nervousness insofar that the hands are being used. Personal conversation with experienced police officers about *modi operandi* that are used for example by thieves, we learned that some of them use tricks like letting something fall right beside the victim to catch his/her attention so that the partner can steal the valuables. Thus, an object formerly used as adaptor might transform into a tool for committing the crime. Therefore, some offenders might carry objects that are useful for committing the crime, but during the build-up phase

they may use them to appear inconspicuously. Paradoxically, exactly this behavior might make them appear suspicious. This would explain the more frequently observed *object-adaptors* during the build-up phase, and also the less frequently observed *self-adaptors* during the build-up phase.

Self-adaptors are also signs of higher arousal that might be expressed more frequently in stress situations (e.g., Sporer & Schwandt, 2007). Thus, it appears also possible that the specific selection of our control condition was responsible for the rather surprising results. Since we used an active search for a (non-existing) person, the "actors" might have felt stressed during the search (control) condition because they did not manage to find the person they were looking for. Some "actors" might have had a very long build-up phase of the criminal act which lead to a longer search (control) condition, because both recordings had the same duration. Therefore, such offenders were observed longer during the search (control) condition and might have expressed more behavior related to stress, because they could not find the non-existing instructor. In this case, the attempted behavioral control factor did not interact with the emotional factor, because the "actors" had no need to hide their intention or to deceive somebody. Thus, the stressful situation might have resulted in the expression of more *self-adaptors*. The combination of stress-related behavior in the search (control) condition, and the interaction of higher arousal and attempted behavioral control during the mock crime condition might have resulted in more frequently shown *self-adaptors* during the search (control) condition.

Similar to Experiment 1, we asked participants in Experiment 2 to also mention which kind of *object-adaptors* and *self-adaptors* they observed in the persons of interest. For *object-adaptors*, the data show a clear pattern: every kind of object-adaptor was significantly more often observed in the mock crime condition than in the search (control) condition. For *self-adaptors*, the pattern appears more unclear. The only significant difference was found for

hands in hair insofar, that this adaptor was more often observed in the mock crime condition. After the results from the overall cluster of behaviors of interest, we would have expected significant differences in the direction of being less often observed in mock crime conditions. A possible reason might be that we controlled for inter-individual differences. Thus, the explicit kinds of behavior (e.g., *hands in face*) might not be expressed that differently, but that the cluster of behavior *self-adaptors* might result in an observable difference.

In general, the results of this pioneering study indicate that offender behavior during the build-up phase of a criminal act indeed is expressed differently compared to non-offender behavior. Along with the importance of controlling for inter-individual differences in future studies, our results also emphasize the high importance of the baseline construct and the ability to successfully establish and adapt a baseline. In other words, observers have to establish the baseline and continuously recheck whether it has changed. During the build-up phase, observers are able to continuously create and update a baseline of nonverbal behavior in the current situational area because bystanders do not have any hidden criminal intentions and therefore show unaltered behavior. This might be an explanation for the contradictory results documented in research on the ability of detecting lies (see Vrij, 2008, for review) and research on the ability of detecting suspicious behavior during the build-up phase (Troscianko et al., 2004; Heubrock et al., 2009a; Baettig et al., 2011). As mentioned before, several differences exist between an interrogation setting and the build-up phase of a criminal act. In interrogation settings, both truth-tellers and liars might alter their behavior (DePaulo et al., 2003; Strömwall et al., 2006) and might feel similar emotions for different reasons (J. K. Burgoon et al., 2009; Vrij, Granhag, & Porter, 2010). This will cause differences due to deception to appear very small only, as can be seen in the small effect sizes of the significant findings in DePaulo et al.'s (2003) meta-analysis. Based on all those differences, it makes sense to exert utmost caution in using nonverbal behavioral cues to deception for lie detection purposes. Detecting criminal intentions during the build-up phase of a criminal act, however,

should be seen as really different for several reasons: Firstly, the tasks of the liar or the criminal before the act differ because a liar directly communicates with the interviewer while the offender during the build-up phase tries to act inconspicuously to slip the security personnel's attention. Secondly, people are able to detect or predict criminal acts in aggressive, unlawful acts (Troscianko et al., 2004), as well as in deceptive, unlawful acts, such as thefts and most terrorist acts (Heubrock et al., 2009a; Baettig et al., 2011; Frey et al., submitted). Thirdly, observers always have an adequate baseline or guideline to judge nonverbal behavioral cues as suspicious or not, which interviewers facing liars are lacking.

Our results show that some nonverbal behavioral cues, which are expressed due to criminal intentions, can indeed be recognized. This is true as long as clusters of behaviors are observed and compared to a baseline. This baseline can either be established from the crowd of bystanders or from the observed person itself in a lawful, comparable situation. We could further show that laypeople indeed show similar moving and communication patterns as professional offenders. Therefore, our results have high ecological validity. We also found nonverbal behavioral cues to suspicious behavior or criminal intentions to be independent of the crime as long as they were compared to a baseline or as long as inter-individual differences are controlled for.

Apart from the limitation that our analyses are based on 13 actors and 23 professional thieves from ten authentic CCTV recordings only, our results are of high practical and theoretical relevance for two reasons. Firstly, they show the importance of creating and using a baseline before judging nonverbal behavior as cues to criminal intentions. Some nonverbal behaviors (e.g., *abrupt direction changes*), which are also mentioned by professionals as cue to criminal intentions, might be more useful and typical for some situational areas than for others. Secondly, our analyses show that there is not just *a* typical nonverbal behavior that can be labeled clearly (e.g., *abrupt changes in direction* or *using a cellphone without instrumental*

goal), but that it is rather a cluster of nonverbal behaviors (e.g., *moving patterns* or *object-adaptors*) that always needs to be compared to an adequate baseline. Therefore, teaching behavior detection officers and police officers possible nonverbal behavioral cues to criminal intentions only does not suffice. They need to know as well what a baseline is, and how it can best be established. As a last step, they need to bring together the observed nonverbal behavioral cues and the baseline. SPOT (U.S. Government Accountability Office, 2010) is a good example that features all these three elements.

Although, there is still more research needed to extract and define more nonverbal behavioral cues to hidden criminal intentions during the build-up phase, this study could show that using behavior detection during the build-up phase as one leg in counterterrorism and preventative police work is a promising step forward.

4 General Discussion

4.1 Summary

The two reported empirical studies filled the lack of scientific research concerning the detection of criminal intentions during the build-up phase of a criminal act. In the first pioneering study (*"Who's the thief?" The influence of knowledge and experience on early detection of criminal intentions*), we examined the influences of theoretical knowledge and work experience on the early detection of hidden criminal intentions. The results suggested that the different stages of expertise, theoretical knowledge and work experience both influence detection performance positively. But, hidden criminal intentions can already be detected without special knowledge at the end of the build-up phase, because at that point even laypeople (i.e. psychology students and police recruits) performed above chance level. Still, inexperienced police officers, having only theoretical knowledge about how some crimes are committed, outperformed laypeople at all points during the build-up phase. The special work experience of criminal investigators then increased performance even more.

To examine what kind of behavior is suspicious (i.e. might help in detecting criminals early on in the build-up phase of their criminal act), we conducted the second reported study (*What is suspicious when trying to be inconspicuous? Criminal intentions inferred from nonverbal behavioral cues*). The results of that study support the belief that hidden criminal intentions lead to nonverbal behavior differing from that of non-suspicious individuals that might trigger suspiciousness. The use of a search task as control condition showed that the suspicious behavior is not mere search behavior due to looking for the perfect victim or place. Rather we found for almost all behaviors of interest (e.g. *moving patterns* in space, *object-* and *self-adaptors*) significant deviation between nonverbal behavior underlying hidden criminal intentions and nonverbal behavior expressed during the search of a person in both

experiments. The only exception concerned *communication behavior*. We assume that the need for communication was similar in all conditions (search, mock crime and real crime) due to the search included in all three tasks. In other words, searching somebody in the search condition might need a similar kind and amount of nonverbal *communication behavior* as during the preparation of a crime, for example, head signs and hand signs. Our results are even more relevant as inter-personal differences were controlled for (i.e. we compared the same people in the mock crime and search condition). Thus, trying to behave unsuspiciously in order to hide a criminal intention indeed alters nonverbal behavior. The first experiment of this study was conducted to examine whether *moving patterns* in space deviate differently from the baseline (how bystanders behave in the same situational area without the same task) when trying to be unsuspicious compared to when somebody is searched for: The deviation from the baseline was less strong for offender conditions than for search conditions. This lends support to the idea of attempted behavioral control from Zuckerman et al.'s (1981) three factor model: offenders might try to become one with the crowd. Furthermore, the results show that, as far as *moving patterns* and *communication behavior* are concerned, our self-made recordings are not different from the authentic CCTV recordings and therefore seem to be comparable to real thefts.

The second experiment of this study was conducted to address *object*- and *self-adaptors* as behaviors of interest. Only self-made recordings were used, because *object*- and *self-adaptors* were not observable in CCTV recordings due to the low quality and distance recording and it should be controlled for inter-personal differences. The results showed a deviation in behavior due to the hidden criminal intention for both behaviors of interest. *Object-adaptors* were shown more strongly in the mock crime condition than in the search condition and *self-adaptors* were shown less strongly in the mock crime condition. Based on the results of the first experiment where offenders seem to try to merge into the crowd (less strong deviation from the baseline than in the search condition), we assume that people might

use *object-adaptors* to appear busy and right at place. On the other hand, they might have the usual belief that *self-adaptors* are signs for nervousness (e.g. Sporer & Schwandt, 2007) and might let them appear suspicious. Thus, they try to control their behavior and show less *self-adaptors* in the mock crime condition than in the search condition.

4.2 Future directions

The central question I addressed in this thesis was whether scientific support to the basic three assumptions underlying an effective behavior detection program can be found: (1) offenders' *hidden* criminal intentions must be displayed in nonverbal behavior that triggers suspiciousness, (2) observers must be able to infer *hidden* criminal intentions based on nonverbal behavioral cues, and (3) the detection of *hidden* criminal intentions during the build-up phase of a criminal act must be teachable and trainable. For all three assumptions the answer is yes - at least to some extent.

Our second study showed promising results with respect to the first assumption: offenders' hidden criminal intentions are expressed in nonverbal behavior, and this nonverbal behavior serves as cues for identification. As the first study indicated, this early identification / detection performance is positively influenced by theoretical knowledge about how a criminal act might be managed. But also (specific) work experience in the field of criminal investigation increases early detection rates. Also the second and third assumptions are met: Early detection of hidden criminal intentions based on nonverbal behavioral cues is possible and it is teachable and trainable, because the different career stages of police officers led to improvements in detection performance.

Even though the second study about diagnostic nonverbal behavioral cues reported here showed promising results, further research is needed. Only four kinds of nonverbal behaviors were addressed, because they were most present in the conversations I had with

criminal investigators. But, behavior detection programs are based on many more behaviors of interest (U.S. Government Accountability Office, 2010) that need further investigation. The first step towards clarifying the usefulness of these nonverbal behavioral cues would be to create good recordings including control for inter-individual differences. Thus, control for possible overestimation or underestimation of cues due to the comparison of different people in the crime and control condition. To increase and/or control ecological validity, more openness between countries concerning CCTV recordings of real terrorist attacks or criminal acts would be welcomed.

In the first empirical study reported in this thesis, criminal investigators outperformed inexperienced police officers and laypeople without being able to label exactly what they are looking for, i.e. which cues they used for their conclusion about suspects. Their excellent performance, however, clearly indicates *that* they consider the relevant cues in the recordings. Therefore, the next step will be to investigate what kind of eye gaze pattern is helpful or needed in correctly identifying deceptive offenders during the build-up phase. Howard, Troscianko, Gilchrist, Behera, and Hogg (2013) provided evidence that trained CCTV operators showed greater between-observer consistency of gaze locations and perceived suspiciousness in dynamic scenes. Thus, CCTV operators seem to know what to look for. The task in this work was to mention whenever something triggers suspiciousness, but there were no real crimes to be observed and therefore no "objective" measure for a good observer. In a study concerning the prediction of antisocial, unlawful acts the visual scan paths of laypeople and CCTV operators did not differ significantly (Grant & Williams, 2011). However, this does not surprise as much as the prediction performance of laypeople and CCTV operators was on chance level for both groups.

In an own attempt we will therefore combine a behavior detection task with eye-tracking measurement and decide afterwards, based on the detection performance, which

persons are good behavior detectors and whose performance was rather poor. Then, in a second step, we will compare the eye-gaze patterns of the good detectors with the eye-gaze patterns of the poor detectors. Difference between eye-gaze patterns might then be taken as indicators for which behavioral cues are diagnostic (e.g. on which parts of the body did they fixate when answering correctly; Grant & Williams, 2011). Moreover, we would like to further investigate whether the poor detectors can learn how to *look correctly* when showing them the scan paths of good detectors (e.g. Neuberger & Körber, 2009).

At the same time, in current behavior detection programs, such as SPOT (Frank et al., 2009; U.S. Government Accountability Office, 2010), observers are trained to look for many different possible diagnostic nonverbal behavioral cues. It is therefore necessary to further investigate how to teach behavior detection best. Behavior detection officers, trained according to the SPOT program, currently learn the concept of a baseline and to make use of it. Additionally, they are taught many diagnostic behavioral cues they have to look for in passengers (U.S. Government Accountability Office, 2010). Other projects, such as the Griffin Project (City of London Police, 2014), are based on raising awareness in all people and show them where to go if they see something that triggers their suspiciousness. For the effectiveness of these programs to be judged adequately it is of crucial relevance to find out whether it is necessary to teach exact and scientifically revealed diagnostic cues, or whether it is enough to teach the baseline concept and raise general awareness in the observers. Based on the second study reported in this thesis, we assume that it is a combination of both, the concept of a baseline and nonverbal behavioral cues. However, based on our results, we emphasize the importance of teaching *clusters* of nonverbal behavioral cues, such as *moving patterns* instead of the exact behavior (e.g. *abrupt changes in directions*). Moreover, the second study revealed that the actual suspicious behavior results in a deviation from the baseline dependent of offender and crime even in both directions. Consequently, we assume the most important part of teaching behavior detection during the build-up phase is teaching

the concept of a baseline and raise awareness that if the taught clusters of behaviors deviate from that baseline in either direction this should raise suspiciousness.

4.3 Conclusion

The pioneering empirical studies reported in this thesis contribute scientific support to the assumption that behavior detection programs might indeed be effective. Early detection of hidden criminal intentions during the build-up phase of a criminal act based on nonverbal behavior is possible. Moreover, the detection performance is increased by theoretical knowledge from police school as well as specific work experience in the field of criminal investigation. Further research is needed to identify the exact processes underlying observers' performance and diagnostic nonverbal behavioral cues to hidden criminal intentions during the build-up phase.

5 References

Anderson, D. E., Ansfield, M. E., & DePaulo, B. M. (1999). Love's best habit: Deception in the context of relationships. In P. Philippot, R. S. Feldman, & E. J. Coats (Eds.), *The social context of nonverbal behavior* (pp. 372–409). Paris, France: Editions de la Maison des Sciences de l'Homme.

Andreassi, J. L. (1995). *Psychophysiology: Human behavior and physiological response* (Vol. 3). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Arsíc, D., Schuller, B., & Rigoll, G. (2007). Suspicious behavior detection in public transport by fusion of low-level video descriptors (pp. 2018–2021). Presented at the IEEE International Conference on Multimedia and Expo, Beijing. doi:10.1109/ICME.2007.4285076

Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology*, *41*(3), 258–290. doi:10.1037/h0055756

Baettig, F., Frey, C., & Hofer, F. (2011). ASPECT (R)-Analysing Suspicious Persons and Cognitive Training Recognition of suspicious Behavior and suspicion Situations. *Kriminalistik*, *65*(10), 641–647.

Baldwin, D. A., & Baird, J. A. (2001). Discerning intentions in dynamic human action. *Trends in Cognitive Sciences*, *5*(4), 171–178. doi:10.1016/S1364-6613(00)01615-6

Bart, E. (2012). *Airport body scanners: The role of advanced imaging technology in airline passenger screening* (p. 12). Congressional Research Service. Retrieved from http://digital.library.unt.edu/ark:/67531/metadc122258/m1/1/high_res_d/R42750_2012Sep20.pdf

BBC News. (2001, September 11). Attacks paralyse New York. *BBC*. Retrieved from http://news.bbc.co.uk/2/hi/americas/1538756.stm

Biggs, A. T., Cain, M. S., Clark, K., Darling, E. F., & Mitroff, S. R. (2013). Assessing visual search performance differences between Transportation Security Administration Officers and nonprofessional visual searchers. *Visual Cognition*, *21*(3), 330–352. doi:10.1080/13506285.2013.790329

Bond, C. F. J., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality & Social Psychology Review (Lawrence Erlbaum Associates)*, *10*(3), 214–234. doi:10.1207/s15327957pspr1003_2

Burgoon, J., Adkins, M., Kruse, J., Jensen, M. L., Meservy, T., Twitchell, D. P., … Younger, R. E. (2005). An approach for intent identification by building on deception detection. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences, 2005. HICSS '05* (pp. 1–10). doi:10.1109/HICSS.2005.78

Burgoon, J. K., & Buller, D. B. (1994). Interpersonal deception: III. Effects of deceit on perceived communication and nonverbal behavior dynamics. *Journal of Nonverbal Behavior*, *18*(2), 155–184. doi:10.1007/BF02170076

Burgoon, J. K., Twitchell, D. P., Jensen, M. L., Meservy, T. O., Adkins, M., Kruse, J., … Younger, R. E. (2009). Detecting concealment of intent in transportation screening: A proof of concept. *IEEE Transactions on Intelligent Transportation Systems*, *10*(1), 103–112. doi:10.1109/TITS.2008.2011700

City of London Police. (2014, January 6). *Project Griffin*. *Project Griffin*. Retrieved July 7, 2014, from http://www.cityoflondon.police.uk/advice-and-support/countering-terrorism/Pages/project-griffin.aspx

Clarke, R. V. G., & Newman, G. R. (2006). *Outsmarting the terrorists*. Greenwood Publishing Group.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorial in Quantitative Methods for Psychology*, *1*(1), 42–45.

Darwin, C. (1872). *The Expression of the Emotions in Man and Animals*. Oxford University Press.

Dente, E., Bharath, A. A., Ng, J., Vrij, A., Mann, S., & Bull, A. (2006). Tracking hand and finger movements for behaviour analysis. *Pattern Recognition Letters*, *27*(15), 1797–1808. doi:10.1016/j.patrec.2006.02.010

DePaulo, B. M. (1992). Nonverbal behavior and self-presentation. *Psychological Bulletin*, *111*(2), 203–243. doi:10.1037/0033-2909.111.2.203

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, *129*(1), 74–118. doi:10.1037/0033-2909.129.1.74

DePaulo, B. M., & Pfeifer, R. L. (1986). On-the-job experience and skill at detecting deception. *Journal of Applied Social Psychology*, *16*(3), 249–267. doi:10.1111/j.1559-1816.1986.tb01138.x

Eachus, P., Stedmon, A., & Baillie, L. (2013). Hostile intent in public crowded spaces: A field study. *Applied Ergonomics*, *44*(5), 703–709. doi:10.1016/j.apergo.2012.05.009

Ekman, P. (1981). Mistakes when deceiving. *Annals of the New York Academy of Sciences*, *364*(1), 269–278. doi:10.1111/j.1749-6632.1981.tb34479.x

Ekman, P. (1993). Facial expression and emotion. *American Psychologist, 48*(4), 384–392. doi:10.1037/0003-066X.48.4.384

Ekman, P. (1998). Afterword: Universality of emotional expressions? A personal history of the dispute. In D. Charles & P. Ekman (Eds.), *The expression of the emotions in man and animals* (3rd ed., pp. 363–393). New York, NY, US: Oxford University Press.

Ekman, P. (2009). *Telling lies: Clues to deceit in the marketplace, politics, and marriage* (4th edition.). New York: Norton & Company.

Ekman, P., & Friesen, W. V. (1969a). Nonverbal leakage and clues to deception. *Psychiatry, 32*, 88–106.

Ekman, P., & Friesen, W. V. (1969b). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica, 1*(1), 49–98.

Ekman, P., & Friesen, W. V. (1972). Hand movements. *Journal of Communication*, *22*(4), 353–374. doi:10.1111/j.1460-2466.1972.tb00163.x

Ekman, P., Friesen, W. V., & O'Sullivan, M. (1988). Smiles when lying. *Journal of Personality and Social Psychology*, *54*(3), 414–420. doi:10.1037/0022-3514.54.3.414

Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., … Tzavaras, A. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology, 53*(4), 712–717. doi:10.1037/0022-3514.53.4.712

Ericsson, K. A. (2008). Deliberate practice and acquisition of expert performance: A general overview. *Academic Emergency Medicine, 15*(11), 988–994. doi:10.1111/j.1553-2712.2008.00227.x

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology*, *47*, 273–305. doi:10.1146/annurev.psych.47.1.273

Frank, M. G., Maccario, C. J., & Govindaraju, V. (2009). Behavior and Security. In P. Seidenstat & F. X. Splane (Eds.), *Protecting airline passengers in the age of terrorism* (pp. 86–106). Santa Barbara, California: ABC-CLIO. Retrieved from http://trid.trb.org/view.aspx?id=919911

Frey, C. I., Wetter, O. E., & Hofer, F. (submitted). "Who's the thief?" The influence of knowledge and experience on early detection of criminal intentions.

Fuhrer, M. (2013). *Die Wirkung von Vorstellungen über Täuschung auf die Wahrnehmung von nonverbalem Verhalten. [The influence of beliefs about deception on the perception of nonverbal behavior.]* (Master Thesis). University of Zurich, Zurich.

Goldstein, E. B. (2007). *Wahrnehmungspsychologie: Der Grundkurs [Sensation and perception, seventh edition]*. (G. Plata, Trans., H. Irtel, Ed.) (7th edition.). Berlin Heidelberg: Spektrum Akademischer Verlag.

Gould, D., Flett, M. R., & Bean, E. (2009). Mental preparation for training and competition. In B. W. Brewer (Ed.), *Handbook of sports medicine and science: Sport psychology* (pp. 53–63). West Sussex, UK: Wiley-Blackwell. Retrieved from http://books.google.ch/books?hl=de&lr=&id=UwI4xd3a5W0C&oi=fnd&pg=PA53&dq=mental+preparation&ots=OBW5PPOrq3&sig=nqouYpvCXwZbPgnU3t2h0keR6Pw

Grant, D., & Williams, D. (2011). The importance of perceiving social contexts when predicting crime and antisocial behaviour in CCTV images. *Legal and Criminological Psychology*, *16*(2), 307–322. doi:10.1348/135532510X512665

Gray, J. A. (1988). *The psychology of fear and stress.* Cambridge University Press.

Green, D. M., & Swets, J. A. (1974). *Signal detection theory and psychophysics* (Vol. 13). Oxford, England: Robert E. Krieger.

Greene, J. O., Dan O'hair, H., Cody, M. J., & Yen, C. (1985). Planning and control of behavior during deception. *Human Communication Research*, *11*(3), 335–364. doi:10.1111/j.1468-2958.1985.tb00051.x

Hardmeier, D., Hofer, F., & Schwaninger, A. (2005). The X-ray object recognition test (X-ray ORT) - a reliable and valid instrument for measuring visual abilities needed in X-ray screening. In *39th Annual 2005 International Carnahan Conference on Security Technology, 2005. CCST '05* (pp. 189–192). doi:10.1109/CCST.2005.1594876

Hardmeier, D., Hofer, F., & Schwaninger, A. (2006). The role of recurrent CBT for increasing aviation security screeners' visual knowledge and abilities needed in X-ray screening. In *Proceedings of the 4th International Aviation Security Technology Symposium* (pp. 338–342). Washington, D.C., USA.

Hartwig, M., Granhag, P. A., Strömwall, L. A., & Vrij, A. (2004). Police officers' lie detection accuracy: Interrogating freely versus observing video. *Police Quarterly*, *7*(4), 429–456. doi:10.1177/1098611104264748

Hecht, M. A., & Ambady, N. (1999). Nonverbal communication and psychology: Past and future. *New Jersey Journal of Communication*, *7*(2), 156–170. doi:10.1080/15456879909367364

Heubrock, D. (2011). Möglichkeiten der polizeilichen Verhaltensanalyse zur Identifikation muslimischer Kofferbomben-Attentäter. [Possibilities of behavior analysis for identifying muslimic suitcase bombers.]. *Polizei-Heute*, *11*(2), 13–24.

Heubrock, D., Immerini, N., Mengeringhausen, U., & Palkies, P. (2009). Das "auffällig unauffällige Verhalten" des Attentäters vor dem Angriff: Zur Psychologie des Personenschutzes [The suspicious unsuspicious behavior of assassins prior to the attack: Psychology of personnel safety]. *Kriminalistik*, (2), 17–23.

Heubrock, D., Kindermann, S., Palkies, P., & Röhrs, A. (2009a). Die Fähigkeit zur Identifikation von Attentätern im öffentlichen Raum: Ein experimenteller Vergleich zwischen professionellen und ungeübten Beobachtern [The ability to identify assassins in public spaces: An experimental comparison between expert and novice observers]. *Polizei & Wissenschaft*, (2), 2–11.

Heubrock, D., Kindermann, S., Palkies, P., & Röhrs, A. (2009b). Verhaltensmerkmale von Attentätern in der Vortatphase: Nonverbale Alarmsignale für Personenschützer. [Behavioral characteristics of assassins during build-up: Nonverbal alarm signals for bodyguards.]. *Polizei & Wissenschaft*, (1), 2–11.

Hofer, F., & Wetter, O. E. (2012). Operational and human factors issues of new airport security technology—two case studies. *Journal of Transportation Security*, *5*(4), 277–291. doi:10.1007/s12198-012-0096-5

Hoogendoorn, S. P., & Bovy, P. H. L. (2004). Pedestrian route-choice and activity scheduling theory and models. *Transportation Research Part B: Methodological*, *38*(2), 169–190. doi:10.1016/S0191-2615(03)00007-9

Howard, C. J., Troscianko, T., Gilchrist, I. D., Behera, A., & Hogg, D. C. (2013). Suspiciousness perception in dynamic scenes: a comparison of CCTV operators and novices. *Frontiers in Human Neuroscience*, *7*, 1–9. doi:10.3389/fnhum.2013.00441

Inbau, F. E., Reid, J. E., Buckley, J. P., & Jayne, B. C. (2001). *Criminal interrogation and confessions* (4th ed.). Fredrick, USA: Aspen Publishers, Inc.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, *14*(2), 201–211. doi:10.3758/BF03212378

Kassin, S. M., & Fong, C. T. (1999). "I'm innocent!': Effects of training on judgments of truth and deception in the interrogation room. *Law and Human Behavior*, *23*(5), 499–516. doi:10.1023/A:1022330011811

Kassin, S. M., Meissner, C. A., & Norwick, R. J. (2005). "I'd know a false confession if I saw one": A comparative study of college students and police investigators. *Law and Human Behavior*, *29*(2), 211–227. doi:10.1007/s10979-005-2416-9

Koivumaki, J. H. (1975). "Body Language Taught Here." *Journal of Communication*, *25*(1), 26–30. doi:10.1111/j.1460-2466.1975.tb00551.x

Kraut, R. (1980). Humans as lie detectors. *Journal of Communication*, *30*(4), 209–218. doi:10.1111/j.1460-2466.1980.tb02030.x

Kuhnke, E. (2008). *Körpersprache für Dummies [Body language for dummies]*. (H. Strahl, Trans.) (Auflage: 1. Auflage.). Weinheim: Wiley-VCH Verlag GmbH & Co. KGaA.

Levine, T. R., Asada, K. J. K., & Park, H. S. (2006). The lying chicken and the gaze avoidant egg: Eye contact, deception, and causal order. *Southern Communication Journal*, *71*(4), 401–411. doi:10.1080/10417940601000576

Mann, S., Vrij, A., & Bull, R. (2004). Detecting True Lies: Police Officers' Ability to Detect Suspects' Lies. *Journal of Applied Psychology*, *89*(1), 137–149. doi:10.1037/0021-9010.89.1.137

Marteau, T. M., & Bekker, H. (1992). The development of a six-item short-form of the state scale of the Spielberger State—Trait Anxiety Inventory (STAI). *British Journal of Clinical Psychology*, *31*(3), 301–306. doi:10.1111/j.2044-8260.1992.tb00997.x

Masip, J. (2005). Is the liar caught sooner than the cripple? Popular wisdom versus scientific knowledge on the non-verbal detection of deception. *Papeles Del Psicologo*, *26*, 78–91.

Masip, J., Alonso, H., Garrido, E., & Antón, C. (2005). Generalized communicative suspicion (GCS) among police officers: Accounting for the investigator bias effect. *Journal of Applied Social Psychology*, *35*(5), 1046–1066. doi:10.1111/j.1559-1816.2005.tb02159.x

McCarley, J. S., Kramer, A. F., Wickens, C. D., Vidoni, E. D., & Boot, W. R. (2004). Visual skills in airport-security screening. *Psychological Science*, *15*(5), 302–306. doi:10.1111/j.0956-7976.2004.00673.x

McCornack, S. A., & Parks, M. R. (1986). Deception detection and relationship development: The other side of trust. *Communication Yearbook*, *9*, 377–389.

Mehrabian, A., & Ferris, S. R. (1967). Inference of attitudes from nonverbal communication in two channels. *Journal of Consulting Psychology*, *31*(3), 248–252. doi:10.1037/h0024648

Mehrabian, A., & Wiener, M. (1966). Non-immediacy between communicator and object of communication in a verbal message: Application to the inference of attitudes. *Journal of Consulting Psychology*, *30*(5), 420–425. doi:10.1037/h0023813

Meissner, C. A., & Kassin, S. M. (2002). "He's guilty!": Investigator bias in judgments of truth and deception. *Law and Human Behavior*, *26*(5), 469–480. doi:10.1023/A:1020278620751

Mitchener-Nissen, T., Bowers, K., & Chetty, K. (2012). Public attitudes to airport security: The case of whole body scanners. *Security Journal*, *25*(3), 229–243. doi:10.1057/sj.2011.20

Mock, T. W. (2009). TSA's new X-ray vision: The fourth amendment implications of body-scan Searches at domestic airport security checkpoints. *Santa Clara Law Review*, *49*, 213.

Morey, R. D. (2008). Confidence intervals from nomalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology*, *4*(2), 61–64.

National Commission on Terrorist Attacks Upon the United States. (2004). *The 9/11 commission report* (pp. 1–586). National Commission on Terrorist Attacks Upon the United States. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a485811.pdf

Neuberger, M., & Körber, B. (2009). *Visuelle Aufmerksamkeitsprozesse im polizelichen Einsatztraining [Visual attention processes in police training]*. Presented at the Polizei Kongress [Police Congress], Austria. Retrieved from http://www.bmi.gv.at/cms/BMI_SIAK/kongress/Vortraege/Neuberger_Koerber_Visuelle_Aufmerksamkeitsprozesse.pdf

Nodine, C. F., Kundel, H. L., Mello-Thoms, C., Weinstein, S. P., Orel, S. G., Sullivan, D. C., & Conant, E. F. (1999). How experience and training influence mammography expertise. *Academic Radiology*, *6*(10), 575–585. doi:10.1016/S1076-6332(99)80252-9

Schwaninger, A. (2005). Increasing efficiency in airport security screeining. *WIT Transactions on the Built Environment*, 407–416.

Schwaninger, A. (2006). Liquid identification: Reacting to the terror threat. Analysing, controlling and adapting to meet new threats. *AIRPORT*, *2006*(5), 30–31.

Schwaninger, A., Hardmeier, D., & Hofer, F. (2004). Measuring visual abilities and visual knowledge of aviation security screeners. In *IEEE ICCST* (Vol. 38, pp. 258–264).

Schwaninger, A., & Hofer, F. (2004). Evaluation of CBT for increasing threat detection performance in X-ray screening. In K. Morgan & M. J. Spector (Eds.), *The Internet Society 2004, Advances in Learning, Commerce and Security.* (pp. 147–156). Wessex: WIT Press.

Schwaninger, A., Hofer, F., & Wetter, O. (2007). Adaptive computer-based training increases on the job performance of X-ray screeners. In *Proceedings of the 41st Carnahan Conference on Security Technology*. Ottawa.

Sporer, S. L., & Schwandt, B. (2007). Moderators of nonverbal indicators of deception: A meta-analytic synthesis. *Psychology, Public Policy, and Law*, *13*(1), 1–34. doi:10.1037/1076-8971.13.1.1

Strömwall, L. A., Hartwig, M., & Granhag, P. A. (2006). To act truthfully: Nonverbal behaviour and strategies during a police interrogation. *Psychology, Crime & Law*, *12*(2), 207–219. doi:10.1080/10683160512331331328

The Global Deception Research Team, T. G. D. R. (2006). A world of lies. *Journal of Cross-Cultural Psychology*, *37*(1), 60–74. doi:10.1177/0022022105282295

Troscianko, T., Holmes, A., Stillman, J., Mirmehdi, M., Wright, D., & Wilson, A. (2004). What happens next? The predictability of natural behaviour viewed through CCTV cameras. *Perception*, *33*(1), 87–101.

U.S. Government Accountability Office. (2010). *Aviation security: Efforts to validate TSA's passenger screening behavior detection program underway, but opportunities exist to*

strengthen validation and address operational challenges (pp. 1–89). Retrieved from http://www.gao.gov/products/GAO-10-763

Von Bastian, C. C., Locher, A., & Ruflin, M. (2013). Tatool: A Java-based open-source programming framework for psychological studies. *Behavior Research Methods*, *45*(1), 108–115. doi:10.3758/s13428-012-0224-y

Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities* (2nd ed., Vol. 13). New York, NY, US: John Wiley & Sons Ltd.

Vrij, A., & Graham, S. (1997). Individual differences between liars and the ability to detect lies. *Expert Evidence*, *5*(4), 144–148. doi:10.1023/A:1008835204584

Vrij, A., Granhag, P. A., & Porter, S. (2010). Pitfalls and opportunities in nonverbal and verbal lie detection. *Psychological Science in the Public Interest*, *11*(3), 89–121. doi:10.1177/1529100610390861

Vrij, A., & Mann, S. (2001). Who killed my relative? Police officers' ability to detect real-life high-stake lies. *Psychology, Crime & Law*, *7*(1-4), 119–132. doi:10.1080/10683160108401791

Wolfe, J. M., Horowitz, T. S., Van Wert, M. J., Kenner, N. M., Place, S. S., & Kibbi, N. (2007). Low target prevalence is a stubborn source of errors in visual search tasks. *Journal of Experimental Psychology: General*, *136*(4), 623–638. doi:10.1037/0096-3445.136.4.623

Younge, G., & Campbell, D. (2003, January 31). Shoe-bomber sentenced to life in prison. Richard Reid restrained by courtroom guards. *The Guardian*. New York, NY, US. Retrieved from http://www.theguardian.com/world/2003/jan/31/usa.uk

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 14, pp. 2–51). New York, NY, US: Academic Press, Inc.

Acknowledgement

First of all, I would like to thank Dr. Franziska Hofer for dragging me into this unique science project and therefore giving me the opportunity to write my thesis in a topic as interesting, exciting and sometimes enervating as this. I am indebted to you for your never ending support, guidance and advice not only concerning my scientific work but also my life. I am deeply grateful for all the things I could learn from your experience in working at the Kantonspolizei Zürich and doing research in an applied area. And I am happily looking forward to further doing research at your side.

My deeply gratitude goes to Prof. Dr. Klaus Oberauer for his openness to this rather unexpected project and therefore for taking me in as a doctoral student. You gave me the unique opportunity to write my thesis. Thank you for your support and advice during the last years, not only concerning scientific work but also concerning cooperation contracts, cooperation partners, as well as writing grant proposals. Thank you for sharing your expertise and experience!

I was very lucky to have Prof. Dr. Claus-Christian Carbon as a second supervisor. Whenever we met; he was a great inspiration and motivation. Thank you for showing me your joy and interest whenever you could.

Additionally, I would like to express my gratitude to Dr. Olive Wetter for sharing his knowledge, experience and most importantly showing me how to write.

The cooperation with the Kantonspolizei Zürich was a striking experience during the past years and I am happy that it will go on. I am deeply thankful to have had the possibility to do a thesis which is important to the law enforcement of Zurich. Special thanks go to Franz Bättig and Felix Walz for being as courageous and innovative as they were for being the first police corps to cooperate with a research institute as deeply as we did. Thank you, Franz

Bättig, for always standing up for the project whenever needed. Another very big thank you goes to Andreas, Hansjörg, Markus, Martin, and Roland (in alphabetical order) for giving me insight into your amazing work as police officers whenever I needed it, thank you for your support and for believing in me and the project. I am still deeply impressed by what you and your colleagues do. Thank you!

Another very big thank you goes to all people from the cognitive psychology lab of the University of Zurich for sharing your expertise and experience with me. I would like to thank Simona Seidmann for your work in the project but also for your friendly support. Thank you, Claudia von Bastian, Carla De Simoni, Mirjam Fuhrer, Laura Hein, Hsuan-Yu Lin, Marcel Niklaus and Mirko Thalmann (in alphabetical order). It was a blast to experience doing a PhD together with you. Michel Druey, thank you for being a mentor and friend during my whole PhD. I would also like to thank especially Mirjam Fuhrer and Tina Subiaz for your humorous texts and phone calls and for your never ending inspiration and motivation throughout writing my thesis, without you I would have struggled even more.

My thesis would certainly not have been possible without the financial support of the Federal Office of Civil Aviation, the Kantonspolizei Zürich, and the University of Zurich, thank you for giving me the opportunity to do my research and complete my thesis.

Last but not least: Without my family and friends, I would certainly not have been able to finish my PhD. Thank you, Mami and Papi, for supporting me throughout my life and give me the strength and possibilities to achieve whatever I wanted; you are world's best parents. Thank you, Nadja, for always believing in me, I am proud to be your sister. Oliver, it is hard to say in words how important you are to me. Thank you for always holding my hand, for loaning me your strength, nerves and belief and for always being there.

Curriculum Vitae of Corinne Ines Frey

Contact	c.frey@psychologie.uzh.ch
Date and Place of Birth	24.06.1983, Sierre, Switzerland

Academic Studies

01/2010-11/2014	PhD student at the Department of Psychology, Cognitive Psychology Unit, University of Zurich Supervisor: Prof. Klaus Oberauer, Co-Supervisor: Prof. Claus-Christian Carbon Doctoral Thesis: "Who's the criminal?" Early detection of criminal intentions - influence of nonverbal behavioral cues, professional experience, and theoretical knowledge Nominated for Mercator Award for Junior Researchers 2014
10/2004-10/2009	MSc at the University of Zurich, Switzerland Psychology (major), Psychopathology of Adolescents and Children and Criminology (minors) Master Thesis: Expertise and Training in Identity Verification

Professional Experience

01/2013-	Research Assistant at the Cognitive Psychology Unit, University of Zurich Project funded by the FOCA and the Zurich State Police
03/2012-12/2012	Research Assistant at the Cognitive Psychology Unit, University of Zurich
2010-	Teacher at the Kantonspolizei Zürich (Zurich State Police) in the Workshop "*ASPECT"*
03/2010-02/2012	Research Assistant at the Kantonspolizei Zürich (Zurich State Police)
07/2008-12/2009	Therapist on behavioral therapy *"ABA"* for autistic children at the KJPD Zurich
07/2007-06/2008	Internship at the KJPD Zurich on behavioral therapy *"ABA"* for autistic children
04/2006 and 08/2006	Internship at the Kindergarten Kosthaus, Schöftland
2006-2008	Student Research Assistant at the Cognitive Psychology Unit, University of Zurich, Prof. W. Marx

Publications

Frey, C. I., Wetter, O. E., & Hofer, F. (submitted). "Who's the thief?" The influence of knowledge and experience on early detection of criminal intentions.

Frey, C. I., Wetter, O. E., & Hofer, F. (submitted). What is suspicious when trying to be inconspicuous? Criminal intentions inferred from nonverbal behavioral cues.

Bättig, F., Frey, C., & Hofer, F. (2011). ASPECT® Analyzing Suspicious Persons and Cognitive Training. Erkennen von verdächtigem Verhalten und verdächtigen Situationen. [ASPECT® Analyzing Suspicious Persons and Cognitive Training. Early detection of suspicious behavior and situations.] *Kriminalistik, 10,* 637-643.

Conference Contributions

Talks:

Frey, C., Chiller-Glaus, S. D., Bättig, F. & Hofer, F. (2011, September). Early Recognition of Suspicious Behavior before Theft. In O. E. Wetter (Chair), *Security research in Switzerland – from X-ray screening to behavior analysis*. Symposium conducted at the meeting of the Swiss Psychological Society, Fribourg (Switzerland).

Posters:

Frey, C., & Hofer, F. (2013, May). *ASPECT® - Wer ist der Dieb? Erkennen verdächtiger Personen anhand ihres nonverbalen Verhaltens. [Who's the thief? Early detection of suspicious people based on nonverbal behavior.]* Poster presented at the LizenziandInnen- Master- und Doktorierenden-Kongress (LiMaDoKo), Zürich (Switzerland).

Frey, C., Bättig, F. & Hofer, F. (2012, March). *ASPECT® - Experience or theoretical knowledge – what is needed to recognize baggage thieves prior to the deed?* Poster presented at the meeting of Tagung experimentell arbeitender Psychologen (TeaP), Mannheim (Germany).

Frey, C., Chiller-Glaus, S. D., Bättig, F. & Hofer, F. (2011, March). *ASPECT® -Früherkennung von auffälligem Verhalten vor einem Diebstahl. [ASPECT® - Early detection of suspicious behavior during the build-up phase of a theft.]* Poster presented at the meeting of Tagung experimentell arbeitender Psychologen (TeaP), Halle (Germany).

Frey, C., Chiller-Glaus, S. D., Bättig, F. & Hofer, F. (2011, May). *ASPECT® -Früherkennung von auffälligem Verhalten vor einem Diebstahl. [ASPECT® - Early detection of suspicious behavior during the build-up phase of a theft.]* Poster presented at the LizenziandInnen- Master- und Doktorierenden-Kongress (LiMaDoKo), Zürich (Switzerland). (3rd prize poster award)

Chiller-Glaus, S. D., & Frey, C. (2008, September). *Is this you? How Training Influences Identification from Photographs.* Poster presented at the LizenziandInnen- und Doktorierenden-Kongress (LiDoKo), Zürich (Switzerland).

Veres-Injac, B., von Bastian, C., Frey, C., Sennhauser, A., & Schwaninger, A. (2007, June). *Matching of Familiar and Unfamiliar Faces with Short Exposure Duration.* Poster presented at the LizenziandInnen- und Doktorierenden-Kongress (LiDoKo), Zürich (Switzerland).